BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 January 2004 (22.01.2004)

PCT

(10) International Publication Number
WO 2004/007670 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/021570

(22) International Filing Date: 10 July 2003 (10.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/395,431 12 July 2002 (12.07.2002) US

(71) Applicant: NEW ENGLAND BIOLABS, INC. [US/US]; 32 Tozer Road, Beverly, MA 01915 (US).

(72) Inventors: MORGAN, Richard, D.; 31 Donovan's Way, Middleton, MA 01949 (US). BHATIA, Tanya; 504 Pinebrook Drive, Peabody, MA 01960 (US). DAVIS, Theodore; 139 Depot Road, Boxford, MA 01921 (US). LOVASCO, Lindsay; 4 Sutton Avenue, Salem, MA 01970 (US).

(74) Common Representative: WILLIAMS, Gregory, D. et al.; New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2004/007670 A2

(54) Title: RECOMBINANT TYPE II RESTRICTION ENDONUCLEASES, MmeI AND RELATED ENDONUCLEASES AND METHODS FOR PRODUCTING THE SAME

(57) Abstract: In accordance with the present invention, there is provided a DNA (deoxyribonucleic acid) fragment which encodes the MmeI type II restriction endonuclease enzyme. This one polypeptide possesses two related enzymatic functions; namely an endonuclease activity which recognizes the DNA sequence 5'-TCC(Pu)AC-3' and cleaves as indicated by the arrows: 5 ' -TCCRAC (N20) ↓-3 ' 3'-AGGYTG(N18)↑-5' and a second enzymatic activity that recognizes the same DNA sequence, 5'-TCC(Pu)AC-3', but modifies this sequence by the addition of a methyl group to prevent cleavage by the MmeI endonuclease activity.

RECOMBINANT TYPE II RESTRICTION ENDONUCLEASES, MmeI AND RELATED ENDONUCLEASES AND METHODS FOR PRODUCING THE SAME

BACKGROUND OF THE INVENTION

The present invention relates to a DNA (deoxyribonucleic acid) fragment, which fragment codes for one polypeptide possessing two related enzymatic functions, namely an enzyme which recognizes the DNA sequence 5'-TCC(Pu)AC-3' and cleaves the phosphodiester bond between the 20th and 21st residues 3' to this recognition sequence on this DNA strand, and between the 18th and 19th residues 5' to the recognition sequence on the complement strand 5'-GT(Py)GGT-3' to produce a 2 base 3' extension (hereinafter referred to as the MmeI restriction endonuclease), and a second enzymatic activity that recognizes the same DNA sequence, 5'-TCC(Pu)AC-3', but modifies this sequence by the addition of a methyl group to prevent cleavage by the MmeI endonuclease. The present invention also relates to a vector containing the DNA fragment, a transformed host containing this DNA fragment, and an improved process for producing MmeI restriction endonuclease from such a transformed host. The present invention also relates to a process for identifying additional DNA fragments that encode enzymes having the same general properties as MmeI but potentially having unique DNA recognition sequences. This process depends on the use of the amino acid sequence of the MmeI enzyme presented in this application, or subsequently on the additional sequences identified through this process. The invention also relates to additional DNA fragments, identifiable through the process described, each of which encodes a polypeptide having significant amino acid sequence similarity to the MmeI polypeptide. The polypeptides

encoded by these DNA fragments are predicted to perform similar functions to MmeI. Specifically, they are predicted to possess the dual enzymatic functions of cleaving DNA in a specific manner at a relatively far distance from the specific recognition sequence and also modifying their recognition sequences to protect the host DNA from cleavage by endonuclease activity. An example of such an enzyme identified by this process is CstMI (see U.S. Application Serial No.:_____, filed concurrently herewith). CstMI was identified as a potential endonuclease because of its highly significant amino acid sequence similarity to MmeI. CstMI recognizes the sequence 5'-AAGGAG-3'and cleaves the phosphodiester bond between the 20th and 21st residues 3' to the recognition sequence on this DNA strand, and between the 18th and 19th residues 5' to the recognition sequence on the complement strand 5'-CTCCTT-3' to produce a 2 base 3' extension.

Restriction endonucleases are a class of enzymes that occur naturally in prokaryotes. There are several classes of restriction systems known, of which the type II endonucleases are the class useful in genetic engineering. When these type II endonucleases are purified away from other contaminating prokarial components, they can be used in the laboratory to break DNA molecules into precise fragments. This property enables DNA molecules to be uniquely identified and to be fractionated into their constituent genes. Restriction endonucleases have proved to be indispensable tools in modern genetic research. They are the biochemical 'scissors' by means of which genetic engineering and analysis is performed.

Restriction endonucleases act by recognizing and binding to particular sequences of nucleotides (the 'recognition sequence') along the DNA molecule. Once bound, the type II endonucleases cleave the molecule within, or to one side of, the sequence. Different restriction endonucleases have affinity for different recognition sequences. The majority of restriction endonucleases recognize sequences of 4 to 6 nucleotides in length, although recently a small number of restriction endonucleases which recognize 7 or 8 uniquely specified nucleotides have been isolated. Most recognition sequences contain a dyad axis of symmetry and in most cases all the nucleotides are uniquely specified. However, some restriction endonucleases have degenerate or relaxed specificities in that they recognize multiple bases at one or more positions in their recognition sequence, and some restriction endonucleases recognize asymmetric sequences. *Hae*III, which recognizes the sequence 5'-GGCC-3', is an example of a restriction endonuclease having a symmetrical, non-degenerate recognition sequence; *Hae*II, which recognizes 5'-(Pu)GCGC(Py)-3' typifies restriction endonucleases having a degenerate or relaxed recognition sequence; while *Bsp*MI, which recognizes 5'-ACCTGC-3' typifies restriction endonucleases having an asymmetric recognition sequence. Type II endonucleases with symmetrical recognition sequences generally cleave symmetrically within or adjacent to the recognition site, while those that recognize asymmetric sequences tend to cleave at a distance of from 1 to 20 nucleotides to one side of the recognition site. The enzyme of this application, MmeI, (along with CstMI) has the distinction of cleaving the DNA at the farthest distance from the recognition sequence of any known type II restriction endonuclease. More than two hundred unique

restriction endonucleases have been identified among several thousands of bacterial species that have been examined to date.

A second component of restriction systems are the modification methylases. These enzymes are complementary to restriction endonucleases and they provide the means by which bacteria are able to protect their own DNA and distinguish it from foreign, infecting DNA. Modification methylases recognize and bind to the same nucleotide recognition sequence as the corresponding restriction endonuclease, but instead of breaking the DNA, they chemically modify one or other of the nucleotides within the sequence by the addition of a methyl group. Following methylation, the recognition sequence is no longer cleaved by the restriction endonuclease. The DNA of a bacterial cell is modified by virtue of the activity of its modification methylase and it is therefore insensitive to the presence of the endogenous restriction endonuclease. It is only unmodified, and therefore identifiably foreign, DNA that is sensitive to restriction endonuclease recognition and cleavage. Modification methyltransferases are usually separate enzymes from their cognate endonuclease partners. In some cases, there is a single polypeptide that possesses both a modification methyltransferase function and an endonuclease function, for example, Eco57I. In such cases, there is a second methyltransferase present as part of the restriction-modification system. In contrast, the MmeI system of the present application has no second methyltransferase accompanying the endonuclease-methyltransferase polypeptide.

Endonucleases are named according to the bacteria from which they are derived. Thus, the species *Haemophilus aegyptius*, for example synthesizes 3 different restriction endonucleases, named *Hae*I, *Hae*II and *Hae*III. These enzymes recognize and cleave the sequences 5'-(W)GGCC(W)-3', 5'-(Pu)GCGC(Py)-3' and 5'-GGCC-3' respectively. *Escherichia coli* RY13, on the other hand, synthesizes only one enzyme, *Eco*RI, which recognizes the sequence 5'-GAATTC-3'.

While not wishing to be bound by theory, it is thought that in nature, restriction endonucleases play a protective role in the welfare of the bacterial cell. They enable bacteria to resist infection by foreign DNA molecules such as viruses and plasmids that would otherwise destroy or parasitize them. They impart resistance by binding to infecting DNA molecules and cleaving them in each place that the recognition sequence occurs. The disintegration that results inactivates many of the infecting genes and renders the DNA susceptible to further degradation by exonucleases.

More than 3000 restriction endonucleases have been isolated from various bacterial strains. Of these, more than 240 recognize unique sequences, while the rest share common recognition specificities. Restriction endonucleases which recognize the same nucleotide sequence are termed "isoschizomers." Although the recognition sequences of isoschizomers are the same, they may vary with respect to site of cleavage (e.g., *Xma*I v. *Sma*I, Endow, et al., *J. Mol. Biol.* **112**:521 (1977); Waalwijk, et al., *Nucleic Acids Res.* **5**:3231 (1978)) and in cleavage rate at various sites (*Xho*I v. *Pae*R7I, Gingeras, et al., *Proc. Natl. Acad. Sci.* U.S.A. **80**:402 (1983)).

Restriction endonucleases have traditionally been classified into three major classes; type I, type II and type III. The type I restriction systems assemble a multi-peptide complex consisting of restriction polypeptide, modification polypeptide, and specificity, or DNA recognition, polypeptide. Type I systems require a divalent cation, ATP and S-adenylosyl-methionine (SAM) as cofactors. Type I systems cleave DNA at random locations up to several thousand basepairs away from their specific recognition site. The type III systems generally recognize an asymmetric DNA sequence and cleave at a specific position 20 to 30 basepairs to one side of the recognition sequence. Such systems require the cofactor ATP in addition to SAM and a divalent cation. The type III systems assemble a complex of endonuclease polypeptide and modification polypeptide that either modifies the DNA at the recognition sequence or cleaves. Type III systems produce partial digestion of the DNA substrate due to this competition between their modification and cleavage activities, and so have not been useful for genetic manipulation.

MmeI does not require ATP for DNA cleavage activity and it cleaves to completion; thus it can be classified as a type II endonuclease. Unlike other type II enzymes, however, MmeI consists of a single polypeptide that combines both endonuclease and modification activities and is sufficient by itself to form the entire restriction modification system. MmeI also cleaves the farthest distance from the specific DNA recognition sequence of any type II endonuclease (as does CstMI of this application). MmeI is quite large and appears to have three functional domains combined in one polypeptide. These consist of an amino-terminal domain which contains the endonuclease DNA cleavage motif and

which may also be involved in DNA recognition, a DNA modification domain most similar to the gamma-class N6mA methyltransferases, and a carboxy-terminal domain presumed to be involved in dimer formation and possibly DNA recognition. The enzyme requires SAM for both cleavage and modification activity. The single MmeI polypeptide is sufficient to modify the plasmid vector carrying the gene *in vivo* to provide protection against MmeI cleavage *in vitro*, yet it is also able to cleave unmodified DNAs *in vitro* when using the endonuclease buffer containing Mg++ and SAM.

There is a continuing need for novel type II restriction endonucleases. Although type II restriction endonucleases which recognize a number of specific nucleotide sequences are currently available, new restriction endonucleases which recognize novel sequences provide greater opportunities and ability for genetic manipulation. Each new unique endonuclease enables scientists to precisely cleave DNA at new positions within the DNA molecule, with all the opportunities this offers.

SUMMARY OF THE INVENTION

In accordance with the present invention, there is provided a novel DNA fragment encoding a novel restriction endonuclease, obtainable from *Methylophilus methylotrophus* (NEB#1190). The endonuclease is hereinafter referred to as "MmeI", which endonuclease:

(1) recognizes the degenerate nucleotide sequence 5'-TCC(Pu)AC-3' in a double-stranded DNA molecule as shown below:

5'-TCC(Pu)AC-3'
3'-AGG(Py)TG-5'

(wherein G represents guanine, C represents cytosine, A represents adenine, T represents thymine, (Pu) represents a purine, either A or G, and (Py) represents a pyrimidine, either C or T);

(2) cleaves DNA in the phosphodiester bond following the 20th nucleotide 3' to the recognition sequence 5'-TCC(Pu)AC-3 and preceding the 18th nucleotide 5' to the complement strand of the recognition sequence 5'-GT(Py)GGA-3' to produce a 2 base 3' extension:

5'-TCC(Pu)AC(N20)/-3'
3'-AGG(Py)GT(N18)/-5'; and

(3) methylates the recognition sequence specified in (1) *in vivo* to protect the host DNA from cleavage by the MmeI endonuclease activity;

The invention further relates to additional DNA fragments, each of which is identified to encode polypeptides which share significant sequence similarity to the MmeI restriction-modification polypeptide. The DNA fragment encoding the MmeI polypeptide enables the identification of these additional potential endonucleases by using similarity searching of the MmeI sequence against sequences available in databases, such as GENBANK, using a program such as BLAST (Altschul, et al. Nucleic Acids Res. 25:3389-3402 (1997)). These DNA fragments, as well as any other fragments with such

similarity to MmeI that may be deposited in the databases in the future, are candidates which may encode polypeptides that are similar to MmeI, in that the polypeptides encoded act as both restriction endonuclease and methyltransferase. These polypeptides may, like MmeI, cleave DNA at a similarly far distance from the recognition sequence, in the range of 18 to 20 nucleotides or more, which character is unique and useful in certain molecular biology technologies. Specifically these polypeptides contain amino acid motifs common to N6mA DNA methyltransferases in the middle of the polypeptide, have a motif common to restriction endonucleases and located in the amino-terminal section of the polypeptides, consisting of the amino acids D/E(X8-X12)D/EXK, and have a region of several hundred amino acids following the conserved methyltransferase motifs which are significantly similar to this region of MmeI and are believed to serve as a dimerization and possibly a DNA sequence recognition domain. An example of such a polypeptide, CstMI, is presented. CstMI has been shown to recognize the 6 base pair asymmetric sequence 5'-AAGGAG-3' and to cleave the DNA in the same manner as MmeI; 5'-AAGGAGN20/N18-3'. The endonuclease encoded by these DNA fragments may be produced by the process used for MmeI, as described below.

The present invention further relates to a process for the production of the restriction endonuclease MmeI. This process comprises culturing a transformed host, such as *E. coli*, containing the DNA fragment encoding the MmeI restriction system polypeptide, collecting the cultured cells, obtaining a cell-free extract therefrom and separating and collecting the restriction endonuclease MmeI from the cell-free extract. The

present invention further relates to a process for the production of the restriction endonucleases encoded by the DNA sequences identified as homologous to MmeI. This process comprises culturing a transformed host, such as *E. coli*, containing the gene for these restriction systems, collecting the cultured cells, obtaining a cell-free extract therefrom and separating and collecting the restriction endonuclease from the cell-free extract.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 - Agarose gel showing MmeI cleavage of lambda, T7, phiX174, pBR322 and pUC19 DNAs.

Figure 2 - DNA sequence of the MmeI gene locus (SEQ ID NO:1).

Figure 3 - Amino acid sequence of the MmeI gene locus (SEQ ID NO:2).

Figure 4 - Agarose gel showing MmeI cleavage of pTBMmeI.1 DNA and unmodified DNA substrates.

Figure 5 - Agarose gel showing MmeI cleavage of unmethylated, hemi-methylated and fully methylated DNA substrates.

Figure 6 - Incorporation of labeled methyl groups into unmethylated, hemi-methylated and fully methylated DNA substrates.

Figure 7 - Multiple sequence alignment of MmeI amino acid sequence (SEQ ID NO:3 through SEQ ID NO:14) and homologous polypeptides from public databases.

DETAILED DESCRIPTION OF THE INVENTION

The recognition sequence and cleavage site of the endonuclease of the present invention were previously described (Boyd, Nucleic Acids Res. 14: 5255-5274 (1986)). However the MmeI enzyme proved difficult to produce from the native host, *Methylophilus methylotrophus*, due to very low yield of the enzyme and the relative difficulty of growing the *M. methylotrophus* host in large quantity. To overcome these limitations to producing MmeI, the present application describes the identification of the DNA sequence encoding the MmeI gene and the expression of this MmeI gene in a suitable host, in the present instance *E. coli*. This manipulation of the MmeI encoding DNA fragment results in both a significant increase in the amount of enzyme produced per gram of cells and a significant increase in ease of growth of large amounts of cells containing MmeI enzyme.

Several standard approaches typically employed by persons skilled in the art of cloning were applied to the task of cloning of MmeI without success. Specifically, the methylase selection approach (Wilson, et al., U.S. Patent No. 5,200,333) was attempted unsuccessfully. Several random libraries of *M. methylotrophus* DNA were constructed in *E. coli* and challenged by digesting with MmeI, but no MmeI methylase containing clones were obtained.

A second approach was also attempted but failed. In this approach, antibodies specific for N6mA were used to screen a library of random clones constructed in a lambda phage replacement vector. The approach was successful in obtaining methylase positive clones, but

all examined were found to express the methyltransferase of the second restriction system in *M. methylotrophus*, the MmeII methylase (recognition sequence 5'-GATC-3') rather than the desired MmeI methylase activity.

The successful approach to obtain the desired DNA fragment encoding the MmeI restriction system involved several steps. First a novel purification procedure was developed to purify the MmeI endonuclease peptide to homogeneity from *M. methylotrophus*. Once this ultra pure MmeI endonuclease polypeptide was successfully obtained in a significant amount, amino acid sequence from the amino terminus and from internal cyanogen-bromide degradation peptides was determined. Using the amino acid sequence obtained, degenerate DNA primers complementary to the DNA coding for the amino acid sequences were synthesized and used to PCR amplify a portion of the MmeI gene. The DNA sequence of this portion of the MmeI gene was determined. The entire MmeI endonuclease gene and surrounding DNA sequences were then obtained by applying the inverse PCR technique. A number of primers matching the DNA sequence obtained were designed, synthesized and used in combination with numerous different templates. The inverse PCR templates were produced by digesting *M. methylotrophus* genomic DNA with various restriction endonucleases and then ligating the cut *M. methylotrophus* DNA at low concentration to obtain circular molecules. The various primers were tried in combinations with the various templates to find primer-template combinations that produced a specific PCR amplification product. The products thus obtained were sequenced. Once the DNA sequence encoding the entire MmeI endonuclease gene was obtained, primers were designed to specifically amplify the gene from *M. methylotrophus* genomic DNA. The amplified gene was

inserted into an expression vector and cloned into *an E. coli* host. The host was tested and found to both express MmeI endonuclease activity and to *in vivo* modify the recombinant expression vector such that it was protected against MmeI endonuclease activity *in vitro*.

This finding that the single polypeptide encoding the MmeI endonuclease also provided *in vivo* protection against MmeI is in contrast to the previously published information on MmeI (Tucholski, Gene 223:293-302 (1998)). Specifically, this reference taught that the MmeI endonuclease polypeptide did not provide protection against MmeI endonuclease cleavage. This reference reported a separate methyltransferase of 48kD as required to modify the MmeI site on both strands and thus block cleavage by the MmeI endonuclease. Specifically, the reference teaches that the MmeI endonuclease polypeptide modifies the adenine in the top strand of the recognition sequence only, 5'-TCCRAC-3' and that such modified DNA is cut by the MmeI endonuclease. The DNA fragment of the present invention encodes the MmeI endonuclease gene, which when grown alone in an *E. coli* host renders the vector containing the MmeI endonuclease resistant to cleavage by the purified MmeI endonuclease. Further, the MmeI endonuclease produced from this fragment does not cleave a DNA fragment modified at the adenine of the top strand, 5'-TCCRAC-3' when no modification of the opposite, or bottom strand is present. This is in contrast to the teaching of the Tucholski reference. Also, the MmeI endonuclease of this application does cleave a DNA fragment in which the adenine residue in the bottom strand is modified 5'-GTYGGA-3' in contrast to the teaching of the Tucholski reference. When both the top strand and the bottom strand are modified at the

adenine residues, the MmeI endonuclease does not cleave the DNA. No second methyltransferase gene, such as reported in the Tucholski reference, was found adjacent to the MmeI endonuclease gene. There is an open reading frame immediately 3' to the MmeI endonuclease gene which would encode a protein of approximately the reported size of such a second methyltransferase activity (48kD). However, this potential polypeptide does not have the amino acid motifs found in methyltransferases, nor did it provide protection against MmeI endonuclease when cloned in *E. coli*. While the Tulchoski reference taught the necessity of a second methyltransferase polypeptide to provide protection against MmeI endonuclease activity for the host cell, it is demonstrated in the present application that the DNA fragment encoding the MmeI endonuclease polypeptide is sufficient to provide such protection. Additionally, the eleven DNA fragments described herein which encode amino acid sequences similar to MmeI are not flanked by any recognizable DNA methyltransferase genes. This indicates that these polypeptides are also likely to provide both protection for the host DNA and endonuclease activity against unmodified DNA substrates on their own, without having a second methyltransferase as part of the restriction modification system. This contrasts with other type II restriction modification systems.

The same group (Tucholski, Gene 223: 293-302 (1998), and Anna Podhajska, personal communication) had previously reported an amino acid sequence of eight residues for a single internal CnBr digestion fragment (sequence GRGRGVGV (SEQ ID NO:___). PCR based on this sequence was attempted yet failed repeatedly. This sequence was found to be unrelated to MmeI once the actual MmeI amino acid sequence was determined in

accordance with the present invention. Therefore correct internal amino acid sequences determination, which enabled the cloning of the MmeI gene, depended on the novel purification method described in this application for the production of sufficiently pure MmeI in large enough quantity to determine cyanogen bromide internal fragment amino acid sequences, as performed in this Application.

In Example II we obtained MmeI by culturing a transformed host carrying the MmeI gene, such as *E. coli* ER2683 carrying pTBMmeI.1 and recovering the endonuclease from the cells. A sample of *E. coli* ER2683 carrying pTBMmeI.1 (NEB#1457) has been deposited under the terms and conditions of the Budapest Treaty with the American Type Culture Collection (ATCC) on July 3, 2002 and bears the Patent Accession No. PTA-4521.

For recovering the enzyme of the present invention *E. coli* carrying pTBMmeI.1 (NEB#1457) may be grown using any suitable technique. For example, *E. coli* carrying pTBMmeI.1 may be grown in Luria broth media containing 100µg/ml ampicillin and incubated aerobically at 37°C with aeration. Cells in the late logarithmic stage of growth are induced by adding 0.3mM IPTG, grown for an additional 4 hours, collected by centrifugation and either disrupted immediately or stored frozen at -70°C.

The MmeI enzyme can be isolated from *E. coli* carrying pTBMmeI.1 cells by conventional protein purification techniques. For example, cell paste is suspended in a buffer solution and treated by sonication, high pressure dispersion or enzymatic digestion to allow extraction of the endonuclease by the buffer solution. Intact cells and cellular debris are

then removed by centrifugation to produce a cell-free extract containing MmeI. The MmeI endonuclease, along with its corresponding intrinsic methylase activity, is then purified from the cell-free extract by ion-exchange chromatography, affinity chromatography, molecular sieve chromatography, or a combination of these methods to produce the endonuclease of the present invention.

The present invention also relates to methods for identifying additional DNA fragments, each of which encodes a polypeptide having significant amino acid sequence similarity to the MmeI polypeptide. The polypeptides encoded by these DNA fragments are predicted to perform similar functions to MmeI. Specifically, they are predicted to possess the dual enzymatic functions of cleaving DNA in a specific manner at a relatively far distance from the specific recognition sequence and also modifying their recognition sequences to protect the host DNA from cleavage by their endonuclease activity. Once the amino acid sequence of the MmeI endonuclease was determined as described in this application, sequences deposited in databases can be compared to this MmeI sequence to find those few sequences that are highly significantly similar to MmeI. This method is similar to that of U.S. Patent No.6,383,770 (Roberts, et al.), except that here we are searching for similarity to the MmeI endonuclease sequence, rather than searching for sequences that match a database of methyltransferase or endonuclease proteins and then examining any unidentified open reading frames next to potential methyltransferase open reading frames. Prior to identifying the MmeI amino acid sequence, the DNA sequences coding for proteins related to MmeI had not been included in the database of restriction and methyltransferase gene sequences utilized by Roberts, et al., *supra* since these sequences had not been linked to

any known endonuclease function. The method disclosed herein of identifying potential MmeI-like endonucleases is thus more specific than the method of U.S. Patent No. 6,383,770 (Roberts, et al.).

Similarity searching of the MmeI sequence against sequences available in databases, such as GENBANK, is accomplished using a program such as BLAST (Altschul, et al. Nucleic Acids Res. 25:3389-3402 (1997)). A sequence with an expectation value (E) score of less than $E= e^{-10}$ is considered a potential candidate endonuclease. Sequences that give expectation values that are much lower, such as less than $E=e^{-30}$ is to be considered as highly likely to be endonucleases like MmeI. Such candidate MmeI-like peptides are further examined to see if they conform to the domain architecture that MmeI exhibits. A true candidate will contain an endonuclease fold motif, usually of the form (D/E)X8-X12(D/E)XK in the amino-terminal portion of the peptide, (Aravind et al. Nucleic Acid Res. 28:3417-3432 (2000)). A true candidate will contain methyltransferase motifs in the middle portion of the peptide similar to gamma class N6-methyl adenine methyltransferases, and sequences similar to the carboxyl portion of MmeI in the carboxyl portion of the candidate peptide. Such a BLAST search performed on June 12, 2003 returned the following sequences as highly significantly similar to MmeI:

Genbank accession ID Description Score E value

GENBANK ACCESSION NO.	DESCRIPTION	SCORE	E VALUE	SEQ ID NO:
1. gi|15794682|ref|NP_284504.1|	hypothetical protein [Neisseri	643	0.0	6
2. gi|9945797|gb|AAG03371.1|	GcrY [Corynebacterium striatum	604	e-171	8
3. gi|16077744|ref|NP_388558.1|	similar to hypothetical protei	564	e-159	7
4. gi|28373198|ref|NP_783835.1|	putative YeeA protein [Lactoba...	531	e-149	3
5. gi|23110638|gb|ZP_00096791.1|	hypothetical protein [Novosph...	426	e-118	10
6. gi|27450519|gb|AAO14619.1|AF465251_62	unknown [Lactobacillus...	217	9e-55	4
7. gi|15807258|ref|NP_295988.1|	DNA modification methyltransfe...	213	1e-53	14
8. gi|15807788|ref|NP_285443.1|	conserved hypothetical protein...	164	7e-39	13
9. gi|21231551|ref|NP_637468.1|	conserved hypothetical protein...	142	2e-32	N/A
10. gi|20803963|emb|CAD31540.1|	PUTATIVE DNA METHYLASE PROTEIN ...	134	7e-30	11
11. gi|23451826|gb|AAN32874.1|AF461726_1	unknown [Pseudomonas f...	98	6e-19	9
12. gi|16125079|ref|NP_419643.1|	conserved hypothetical protein...	92	3e-17	12
13. gi|10954534|ref|NP_044172.1|	M. jannaschii predicted coding...	76	2e-12	N/A

Most of these proteins are labeled as hypothetical or putative in their database entries. A number of these appear to be full-length polypeptides, such as sequence #2 above: GcrY. Such candidates can be expressed as described in Roberts to identify the expected endonuclease activity. Some endonuclease genes may be inactive in the particular strain used for sequencing (Lin, et al. Proc. Natl. Acad. Sci. USA 98:2740-2745 (2001)). In such a circumstance it may prove possible to express functional endonucleases by repairing the mutations that have inactivated these genes. Several of the MmeI homologs, such as #7 (SEQ ID NO:14)(Deinococcus radiodurans DR2267) and #8 (SEQ ID NO:13)(Deinococcus radiodurans DR0119.1) have disruptions in the open reading frames. DR2267 has a stop codon, TAG, which prematurely terminates the open reading frame, in a position where MmeI has a glutamate amino acid coded for by the codon GAG. By changing this TAG stop codon to GAG it may be possible to reactivate this potential endonuclease gene. DR0119.1 is also disrupted, in that it has a frameshift that disrupts open reading frame. The MmeI sequence may be used as a guide to direct where to repair this frameshift by maximizing the similarity of the DR0119.1 sequence to the MmeI sequence. This may well restore DR0119.1 endonuclease activity.

An alternative way to generate potential new endonucleases is to take advantage of their similar domain structure by performing domain swapping. One may be able to swap the amino terminal domain of an MmeI-like peptide, for the amino terminal domain in the MmeI protein, for example by swapping the sequence of the potential new gene up to the first methyltransferase motif (motif X, "Gly Ala His Tyr Thr Ser" into MmeI to replace this portion of MmeI up to the same sequence.

This approach may be particularly useful when only a partial sequence is available or a potential gene has lost function due to multiple mutations. This approach will create a chimeric protein that potentially has endonuclease activity and cleaves at a distance away from the recognition sequence, like MmeI, but that recognizes a novel DNA sequence. One may also find sequences in the databases that are highly similar to MmeI but that are partial. For example, sequence #11 (SEQ ID NO:9) above (Pseudomonas fluorescens) is from a small fragment of DNA sequence in the database. To obtain a functional endonuclease like MmeI from this sequence one can use inverse PCR or other techniques to obtain DNA sequence adjacent to the fragment reported, then use that sequence to obtain an intact endonuclease gene.

Once a sequence is identified the potential endonuclease can be expressed and characterized as described in Roberts, et al. *supra*. Here, however, there is no separate methyltransferase gene to express along with the endonuclease. Once such a potential endonuclease is cloned and expressed in a suitable host, such as in E. coli, a cell free extract is prepared and analyzed to detect any endonuclease activity. Such an endonuclease assay must include the SAM cofactor required by these endonucleases. Once specific DNA cleavage activity is found the recognition sequence and cleavage site may be determined by standard methods. (Schildkraut, (1984) In Genet. Eng. (N Y) Vol 6. (Setlow J.K., Hollaender, A. Ed.). pp 117-140. Plenum Press, New York. "Screening for and characterizing restriction endonucleases.")

The enzymes so identified can be isolated from *E. coli* cells carrying the DNA fragment in a suitable vector by conventional protein purification techniques. For example, cell paste is suspended in a buffer solution and treated by sonication, high pressure dispersion or enzymatic digestion to allow extraction of the endonuclease by the buffer solution. Intact cells and cellular debris are then removed by centrifugation to produce a cell-free extract containing the enzyme. The endonuclease, along with its corresponding intrinsic methylase activity, is then purified from the cell-free extract by ion-exchange chromatography, affinity chromatography, molecular sieve chromatography, or a combination of these methods to produce the endonuclease of the present invention.

These DNA fragments, as well as any other fragments with such similarity to MmeI that may be deposited in the databases in the future, are predicted to encode polypeptides that are similar to MmeI, in that the polypeptides encoded act as both restriction endonuclease and methyltransferase. These polypeptides may, like MmeI, cleave DNA at a similarly far distance from the recognition sequence, in the range of about 18 to 20 nucleotides or more, which character is unique and useful in certain molecular biology technologies.

An example of such an enzyme identified by this process is CstMI (see U.S. Application Serial No._____, filed concurrently herewith). CstMI was identified as a potential endonuclease because of its highly significant amino acid sequence similarity to MmeI. CstMI is encoded by sequence #2 above (SEQ ID NO:8), which gave highly significant Expectation value of e^{-171} when compared to MmeI by BLAST. CstMI recognizes

the 6 base pair asymmetric sequence 5'-AAGGAG-3' and cleaves the DNA in the same manner as MmeI: it cleaves the phosphodiester bond between the 20th and 21st residues 3' to this recognition sequence on this DNA strand, and between the 18th and 19th residues 5' to the recognition sequence on the complement strand 5'-CTCCTT-3' to produce a 2 base 3' extension.

The present invention is further illustrated by the following Examples. These Examples are provided to aid in the understanding of the invention and are not construed as a limitation thereof.

The references cited above and below are herein incorporated by reference.

EXAMPLE I

PURIFICATION OF *Mme*I ENDONUCLEASE

A single colony of *Methylophilus methylotrophus* (NEB#1190) was grown for 24 hrs in 1 liter of medium M (0.08 μM $CuSO_4$, 0.448 μM $MnSO_4$, 0.348 μM $ZnSO_4$, 6.0 μM $FeCl_3$, 18 μM $CaCO_3$, 1.6 mM $MgSO_4$, 9.0 mM NaH_2PO_4, 10.9 mM K_2HPO_4, 13.6 mM $(NH_4)_2SO_4$) for 24 hours. This culture was used to inoculate 100 liters of medium M. The cells were grown aerobically at 37°C, overnight, until stationary. Five 100-liter fermentations were required to harvest 752 grams of wet cell pellet.

750 gram of *M. methylotrophus* cell pellet was suspended in 2.25 liters of Buffer A (20 mM Tris-HCl (pH 8.0), 50 mM NaCl, 1.0 mM DTT, 0.1 mM EDTA, 5% Gycerol) and passed through a Gaulin homogenizer at ~12,000 psig.

The lysate was centrifuged at ~13,000 x G for 40 minutes and the supernatant collected.

The supernatant solution was applied to a 500 ml Heparin Hyper-D column (BioSepra SA) which had been equilibrated in buffer A. A 1.0 L wash of buffer A was applied, then a 2 L gradient of NaCl from 0.05 M to 1 M in buffer A was applied and fractions were collected. Fractions were assayed for Mme I endonuclease activity by incubating with 1 μg Lambda DNA (NEB) in 50 μl NEBuffer 1, supplemented with 32 μM S-adenosyl-L-methionine (SAM) for 15 minutes at 37° C. MmeI activity eluted at 0.3 M to 0.4 M NaCl.

The Heparin Hyper-D column fractions containing the Mme I activity were pooled, diluted to 50 mM NaCl with buffer A (without NaCl) and applied to a 105 ml Source15 Q column (Amersham Biotech) which had been equilibrated with buffer A. A 210 ml wash with buffer A was applied followed by a 1.0 L gradient of NaCl from 0.05 M to 0.7 M in buffer A. Fractions were collected and assayed from Mme I endonuclease activity. The Mme I activity was found in the unbound fraction.

The Source15 Q pool was loaded onto a 22 ml AF-Heparin-TSK column (TosoHaas) which had been equilibrated with buffer A. A wash of 44 ml buffer A was applied, followed by a linear gradient of NaCl from 0.05 M to 1.0 M in buffer A. Fractions were collected and assayed from Mme I endonuclease activity. The Mme I activity eluted between 0.26 M and 0.29 M NaCl. The fractions containing activity were pooled and dialyzed against buffer B (20 mM $NaPO_4$ (pH 7.0), 50 mM NaCl, 1.0 mM DTT, 0.1 mM EDTA, 5% Glycerol).

The dialyzed AF-Heparin-TSK pool was loaded onto a 6 ml Resource15 S column (Amersham Biotech) which had been equilibrated with buffer B. A wash of 12 ml buffer B was applied, followed by a linear gradient of NaCl from 0.05 M to 1.0 M in buffer B. Fractions were collected and assayed for Mme I endonuclease activity. Mme I activity eluted between 0.14 M and 0.17 M NaCl.

This pool was applied to a 2 liter Superdex 75 sizing column (Amersham Biotech) which had been equilibrated with buffer C (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1.0 mM DTT, 0.1 mM EDTA, 5% Glycerol). Fractions were collected between 500 and 1500 ml elution with buffer C, then assayed by Mme endonuclease assay and polyacrylamide gel electrophoresis on 4-20% gradient gel, followed by protein staining with Coomassie Brilliant Blue dye. Fractions eluting between 775 and 825 ml corresponded to Mme I activity and a protein band of 105 kDa. These fractions were pooled and dialyzed against buffer D (20 mM $NaPO_4$ (pH 7.0), 50 mM NaCl, 1 mM DTT, 5% Glycerol).

The dialyzed sizing pool was applied to a 16 ml Ceramic HTP column (BioRad) which had been equilibrated with buffer D. A 32 ml wash with buffer D was followed by a linear gradient from 0.02 M to 1.0 M $NaPO_4$ in buffer D. Fractions were collected and assayed by Mme endonuclease assay and polyacrylamide gel electrophoresis on a 4-20% gradient gel, followed by protein staining with Coomassie Brilliant Blue dye. Mme I eluted between 0.26 M and 0.3 M $NaPO_4$. A portion of several fractions containing a single homogeneous protein band of 105 kDa were used for protein sequencing. The rest of the purified MmeI fractions were pooled (6 ml @ .36 mg/ml) and dialyzed against storage

buffer (10 mM Tris (pH 7.9), 50 mM KCl, 1mM DTT, .1 mM EDTA, 50% glycerol). The purified MmeI enzyme was stored at -20°C.

Activity determination:

Samples from 1-4 µl were added to 50 µl substrate solution consisting of 1X NEBuffer 1, 32 µM S-adenosyl-L-methionine, and 1 µg DNA (lambda, PhiX174 or pUC19 DNAs). Reactions were incubated for 15 minutes at 37°, received 20 µl stop solution and were analyzed by electrophoresis on a 1% agarose gel.

Optimized endonuclease activity

Following purification of MmeI from *M. methylotrophus,* experiments were performed to determine the optimal reaction conditions for DNA cleavage. Endonuclease activity was found to be significantly enhanced by the presence of potassium in the reaction buffer. Reactions were performed at 4°C to 37°C and from 5 to 60 minutes with no appreciable change in the amount of DNA cleavage. Enzyme concentrations at or near stoichiometric equivalence to DNA sites were required for maximal cleavage. Large excess of enzyme blocked cleavage. These findings were used to reassess the activity of MmeI and to define a workable endonuclease unit.

Unit definition

One unit of MmeI is defined as the amount of MmeI required to completely cleave 1 µg of PhiX174 DNA in 15 minutes at 37°C in NEBuffer 4 (20 mM Tris-acetate, 10 mM magnesium acetate, 50 mM potassium acetate, 1 mM

dithiothreitol (pH7.9 at 25°C)) supplemented with 80μM S-adenosyl-L-methionine (SAM).

EXAMPLE II

CLONING THE *MmeI* ENDONUCLEASE

1. DNA purification: Total genomic DNA of *Methylophilus methylotrophus* was prepared. 5 grams of cell paste was suspended in 20 ml of 25% sucrose, 0.05 M Tris-HCl pH 8.0, to which was added 10 ml of 0.25 M EDTA, pH 8.0. Then 6 ml of lysozyme solution (10 mg/ml lysozyme in 0.25 M Tris-HCl, pH 8.0) was added and the cell suspension was incubated at 4°C for 16 hours. 25 ml of Lytic mix (1% Triton-X100, 0.05 M Tris, 62 mM EDTA, pH 8.0) and 5 ml of 10% SDS was then added and the solution incubated at 37°C for 5 minutes. The solution was extracted with one volume of equilibrated phenol:chloroform:isoamyl alcohol (50:48:2, v/v/v) and the aqueous phase was recovered and extracted with one volume of chloroform:isoamyl alcohol (24:1, v/v) two times. The aqueous solution was then dialysed against four changes of 2 L of 10 mM Tris, 1 mM EDTA, pH 8.0. The dialysed DNA solution was digested with RNase (100 μg/ml) at 37°C for 1 hour. The DNA was precipitated by the addition of 1/10th volume 5 M NaCl and 0.55 volumes of 2-propanol and spooled on a glass rod. The DNA was briefly rinsed in 70% ethanol, briefly air dried and dissolved in 20 ml TE (10 mM Tris, 1 mM EDTA, pH 8.0) to a concentration of approximately 500 μg/ml and stored at 4°C.

2. The MmeI endonuclease was purified to homogeneity as described in Example I above.

3. Amino acid sequences of the *Mme*I endonuclease were obtained for the amino terminus and for several internal cyanogen bromide digestion products of the MmeI polypeptide. The MmeI restriction endonuclease, prepared as described in Example I above, was subjected to electrophoresis and electroblotted according to the procedure of Matsudaira (Matsudaira. J. Biol. Chem. 262:10035-10038, 1987)), with modifications as previously described (Looney, et al. Gene 80:193-208 (1989)). The membrane was stained with Coomassie blue R-250 and the protein band of approximately 105 kD was excised and subjected to sequential degradation on an ABI Procise 494 Protein/Peptide Sequencer with gas-phase delivery (Waite-Rees, et al. J. Bacteriol. 173:5207-5219 (1991)). The amino acid sequence of the first 14 amino terminal residues obtained was the following: ALSWNEIRRKAIEF (SEQ ID NO:15).

An additional sample of the MmeI endonuclease, 20 μg in 20 μl, was treated with 2 μg of cyanogen bromide (Sigma) dissolved in 200 μl of 88% distilled formic acid for 24 hours in the dark at room temperature. This reaction mixture was evaporated to dryness and resuspended in 20 μl of loading buffer (1.5M Tris-HCl, pH 8.5, 12% glycerol, 4% SDS, 0.05% Serva Blue G, 0.05% Phenol Red) at 100°C for 5 minutes. This sample was subjected to electrophoresis on a Tris-Tricine 10 to 20% polyacrylamide gradient gel (Invitrogen) for three hours and then transferred to a polyvinylidene difluoride (PVDF) membrane (Problott, Applied Biosystems Inc.) using 10 mM CAPS buffer (10mM 3-[cyclohexylamino]-1-propanesulfonic acid, 10% methanol, 0.05% SDS, 0.005% dithiotheritiol, adjusted to pH 11.0 with NaOH) for 18 hours at 200 volts in a tank electroblotter (TE52, Hoeffer). The membrane was stained with Coomassie blue

R-250 and major bands of 25 kilodaltons (kD), 14 kD, 7.5 kD and 6 kD were observed, as well as smaller bands. These stained protein bands were excised from the membrane and each subjected to sequential degradation. The fragments other than the amino terminal fragment are derived from internal cleavage by cyanogen bromide at methionine residues from within the protein and thus should be preceded by a methionine. The first 29 residues of the 25 kD peptide corresponded to (M)KISDEFGNYFARIPLKSTXXIXEXNALQ (SEQ ID NO:16). Residues 20, 21, 23 and 25, labeled X, were not identified. The first 40 amino acid residues obtained from the 14kD fragment were: (M)DAKKRRNLGAHYTSEANILKLI KPLLLDELWVVFXKVKN (SEQ ID NO:17). Residue 36 was not determined. The first 25 residues of the 7.5 kD peptide corresponded to (M)KSRGKDLDKAYDQALDYFSGIAER (SEQ ID NO:18). The 6kD fragment was found to contain a mixture of three sequences.

4. Amplification of a portion of the MmeI endonuclease: The peptide sequence data from the amino terminus, 25 kD, 14kD and 7.5kD peptides was used to construct a series of degenerate PCR primers corresponding to the codons for the amino acid residues. The order of the internal peptide fragments was unknown, so both forward (sense strand) and reverse (antisense strand) primers were made for these fragments. The primers were:

25 kD fragment: residues DEFGNYFA (SEQ ID NO:19)

Forward:

1) 5'-GARTTYGGNAAYTAYTTYGC-3' (SEQ ID NO:20)

Reverse:

2) 5'-AARTARTTNCCRAAYTCRTC-3' (SEQ ID NO:21)

14 kD fragment: residues MDAKKR (SEQ ID NO:22)

Forward A:

3) 5'-ATGGAYGCNAARAARCG-3' (SEQ ID NO:23)

Forward B:

4) 5'-ATGGAYGCNAARAARAG-3' (SEQ ID NO:24)

Reverse:

5) 5'-CGNCGYTTYTTNGCRTCCAT-3' (SEQ ID NO:25)

7.5 kD fragment: residues DKAYDQA (SEQ ID NO:26)

Forward:

6) 5'-GAYAARGCNTAYGAYCARGC-3' (SEQ ID NO:27)

Reverse:

7) 5'-GCYTGRTCRTANGCYTTRTC-3' (SEQ ID NO:28)

where
Y = T,C
R = A,G
H = A,T,C
S = G,C
N = A,C,G,T

Primers 1 and 2 are derived from the MmeI 25 kD CNBr peptide and were prepared to prime on the sense strand (1) or the antisense strand (2) of the gene. Primers 3 through 5 are derived from the 14 kD CNBr peptide and were prepared to prime on the sense strand (3 and 4) or the antisense strand (5) of the gene, with 3 and 4 differing in the codon usage for the arginine residue. Primers 6 and 7 are derived from the 7.5 kD CNBr peptide and were prepared to prime on the sense strand (6) or the antisense strand (7) of the gene.

PCR amplification reactions were performed using the primer combinations of 1 with 5, 1 with 7, 3 with 2,

3 with 7, 4 with 2, 4 with 7, 6 with 2 and 6 with 7. A portion of the MmeI gene was amplified in a PCR reaction by combining:

80 µl 10X Thermopol buffer (NEB)
50 µl 4mM dNTP solution (NEB)
4 µl MmeI genomic DNA (500µg/ml stock)
16 µl 100mM $MgSO_4$
586 µl dH_2O
16 µl (32 units) Vent® exo- DNA polymerase (NEB).

This master mix was divided into 8 aliquots of 90 µl, to which was added 5 µl forward primer (10 µM stock) and 5 µl reverse primer (10 µM stock). The cycling parameters were 95°C for 3 minutes for one cycle, then 95°C for 30 seconds, 46°C for 30 seconds, 72°C for 2 minutes, for 25 cycles.

The amplification reactions were electrophoresed on a 1% agarose gel and analyzed. Major DNA amplification products of 450 base pairs (bp) (primers 2 with 4), 650 bp (primers 5 with 6) and 1100 bp (primers 2 with 6) were obtained. These fragment sizes are consistent with the 7.5 kD CnBr fragment being located nearest the amino end of the protein and approximately 650 bp away from the 14kD CnBr fragment, with the 14 kD fragment between the 7.5 kD and the 25 kD fragment and adjacent to the 25 kD fragment. The amplified DNA fragments were gel purified and sequenced using the primers that were used for the amplification. A translation of the DNA sequence obtained matched the amino acid sequence derived from the purified MmeI endonuclease, indicating that a portion of the MmeI endonuclease gene DNA sequence had been successfully obtained.

5. Determining the DNA sequence for the entire MmeI gene and adjacent DNA: The inverse PCR technique was used to extend the DNA sequence from both sides of the 1060 bp of the MmeI gene obtained above. To accomplish this a series of primers matching the MmeI gene DNA sequence and oriented for inverse PCR were designed and synthesized. MmeI genomic DNA was cut with a number of restriction endonucleases and ligated at low concentration to generate circular DNA templates.

A. MmeI genomic DNA was digested with ten different restriction endonucleases and then circularly ligated to obtain DNA templates to amplify using the inverse PCR technique. The restriction enzymes used were:

BspHI (T/CATGA)
EcoRI (G/AATTC)
HindIII (A/AGCTT)
HinP1I (G/CGC)
MspI (C/CGG)
NlaIII (CATG/)
PstI (CTGCA/G)
SacI (GAGCT/C)
SphI (GCATG/C)
XbaI (T/CTAGA)

Restriction enzyme digests were performed by combining:

5 µl 10X NEBuffer recommended for the enzyme (varied with enzyme)
2 µl *M. methyloptrophus* genomic DNA (1 µg)
43 µl dH_2O
1 µl (10 - 20 units) restriction enzyme.

The reactions were incubated for 1 hour at 37°C. The restriction endonuclease was inactivated by heating

the reaction to 65°C (80°C for PstI) for 20 minutes. The digested DNA was then ligated into circular fragments by adding 50 µl 10X T4 DNA ligase buffer, 400 µl dH_2O and 3 µl concentrated T4 DNA ligase (6000 units, New England Biolabs, Inc.) and incubating at 16°C for 16 hours. The ligated DNA was then extracted with phenol and chloroform, precipitated with 2-propanol and resuspended in 100 µl TE buffer.

B. Amplification of DNA adjacent to the 1060 bp fragment of the MmeI endonuclease gene: Two pairs of PCR primers were designed, one near each end of the 1060 bp sequence obtained from direct PCR with degenerate primers. The primer sequences were:

primer IP 1:

5'-GTTGGATCCCGCACAGATTGCTCAGG-3' (SEQ ID NO:29)

primer IP 2:

5'-GTTGGATCCTACGTTAATCTGAATAAGATG-3' (SEQ ID NO:30)

primer IP 3:

5'-GTTGGATCCTGTTAATCTGAAACGCTGG-3' (SEQ ID NO:31)

primer IP 4:

5'-GTTGGATCCTTATACCAAAATGTGAGGTC-3' (SEQ ID NO:32)

Inverse PCR reactions were performed on the 10 circularized templates produced above with the primer pairs of IP 1 with IP 2, IP 3 with IP 4, and IP 1 with IP 3. The amplification reactions were assembled by combining:

80 µl 10X Thermopol buffer (NEB)
50 µl 4mM dNTP solution (NEB)
40 µl IP primer (forward)

40 µl IP primer (reverse)
16 µl 100mM $MgSO_4$
534 µl dH_2O
16 µl (32 units) Vent® exo- DNA polymerase (NEB).

The master mix was aliquoted into ten tubes of 76 µl, to which was added 4 µl of the appropriate digested, circularly ligated template. The cycling parameters were 95°C for 3 minutes for one cycle, then 95°C for 30 seconds, 56°C for 30 seconds, 72°C for 3 minutes, for 25 cycles. Amplification products were analyzed by agarose gel electrophoresis.

For primers IP 1 and IP 2 with the SphI template and the NlaIII template a product of approximately 825 bp was obtained. For primers IP 3 and IP 4 with the BspHI template a product of approximately 800 bp was obtained. For primers IP 1 and IP 3 with the EcoRI template a product of approximately 1500 bp was obtained. These amplified DNA fragments were gel purified, sequenced and assembled with that previously obtained. The assembled sequence did not contain the entire MmeI endonuclease open reading frame. The assembled sequence was used to direct synthesis of a second group of inverse PCR primer pairs. The sequences of these primers were:

primer IP 5:

5'-TTCAGAAATACGAGCGATGC-3'(SEQ ID NO:33)

primer IP 6:

5'-GTCAAGCCATAAACACCATC-3'(SEQ ID NO:34)

primer IP 7:

5'-GAGGGTCAGAAAGGAAGCTG-3'(SEQ ID NO:35)

primer IP 8:

5'-GTCCAACTAACCCTTTATGG-3'(SEQ ID NO:36)

Inverse PCR amplification reactions were performed as above. Using primers IP 5 and IP 6, products were obtained from the NlaIII template (approximately 450 bp) and the MspI template (approximately 725 bp), but not from the other circular ligation templates. Using primers IP 7 and IP 8, products were obtained from the EcoRI template (approximately 500 bp), the SphI template (approximately 825 bp) and the BspHI template (approximately 750 bp). These DNA fragments were sequenced and the sequence was assembled with that previously obtained. The assembled sequence did not yet contain the entire MmeI endonuclease open reading frame, so another round of primer synthesis and inverse PCR was performed. Additional DNA templates were generated as above, but using the restriction enzymes ApoI (R/AATTY), AseI (AT/TAAT), BsaHI (GR/CGYC), MfeI (C/AATTG), SspI (AAT/ATT) and EcoRV (GAT/ATC) to digest *M. methylotrophus* genomic DNA. The sequences of this third round of primers were:

primer IP 9:

5'-TTCCTAGTGCTGAACCTTTG-3' (SEQ ID NO:37)

primer IP 10:

5'-GTTGCGTTACTTGAAATGAC-3' (SEQ ID NO:38)

primer IP 11:

5'-CCAAAATGGAACTTGTTTCG-3' (SEQ ID NO:39)

primer IP 12:

5'-GTGAGTGCGCCCTGAATTAG-3' (SEQ ID NO:40)

Inverse PCR amplification reactions were performed as above. Using primers IP 9 and IP 10, products were obtained from the NlaIII template (approximately 425 bp), the MfeI template (approximately 750 bp), the ApoI template (approximately 800 bp) and the MspI template (approximately 2100 bp). Using primers IP 11 and IP 12, products were obtained from the SphI template (approximately 875 bp), the BspHI template (approximately 925 bp) and the EcoRI template (approximately 950 bp). These DNA fragments were sequenced and the sequence was assembled with the sequences previously obtained. Further sequencing was performed on the IP 9, IP10 MspI 2100 bp product using three additional primers:

primer S1:

5'-GCTTCATTTCATCCTCTGTGC-3' (SEQ ID NO:41)

primer S2:
5'-TAACCGCCAAAATTAATCGTG-3' (SEQ ID NO:42)

primer S3:

5'-CCACTATTCATTACAACACC-3' (SEQ ID NO:43)

The final assembled sequence (Figure 2) contained the entire MmeI restriction gene, as well as 1640 bp of sequence preceding the gene and 1610 bp of sequence following the gene.

6. Cloning the MmeI endonuclease gene in *E. coli*: The putative MmeI endonuclease open reading frame was identified from the DNA sequence assembly obtained

from sequencing the various inverse PCR amplified DNA fragments. The beginning of the open reading frame was identified on the basis of the match of the predicted amino acid sequence at the amino terminus of the open reading frame with the sequence determined from the MmeI endonuclease protein. The predicted end of the open reading frame would allow for the coding of an approximately 105 kD polypeptide, which matched the observed size of the native MmeI endonuclease. The amino acid sequence deduced from translation of this open reading frame contained conserved sequence motifs of N6mA DNA methyltransferases. However, no open reading frame containing sequence motifs conserved among DNA methyltransferases was observed adjacent to the MmeI endonuclease gene, as had been expected. It was decided to try to express the MmeI endonuclease in *E. coli* without having a second methyltransferase present to protect the *E. coli* host DNA from cleavage. Oligonucleotide primers were synthesized to specifically amplify the MmeI gene from *M. methylotrophus* genomic DNA for expression in the cloning vector pRRS (Skoglund, Gene 88:1-5 (1990)). The forward primer contained a PstI site for cloning, a stop codon in frame with the lacZ gene of the vector, a consensus E. coli ribosome binding site, the ATG start codon for translation (changed from the GTG used by *M. methylotrophus* to facilitate greater expression in E. coli) and 20 nucleotides that matched the *M. methylotrophus* DNA sequence:

5'-GTTCTGCAGTTAAGGATAACATATGGCTTTAAGCTGGAACGAG-3'
(SEQ ID NO:44)

The reverse primer contained a BamHI site for cloning and 22 nucleotides that matched the *M.*

methylotrophus DNA sequence 3' to the end of the MmeI open reading frame:

5'-GTTGGATCCGTCGACATTAATTAATTTTTGCCCTTAG-3'
(SEQ ID NO:45)

The MmeI gene was amplified in a PCR reaction by combining:

50 µl 10X Thermopol buffer (NEB)
30 µl 4mM dNTP solution
12.5 µl forward primer (10µM stock)
12.5 µl reverse primer (10µM stock)
5 µl MmeI genomic DNA (500µg/ml stock)
387 µl dH_2O
3 µl (6 units) Vent® DNA polymerase

The reaction was mixed and aliquoted into 5 tubes of 80 µl each. $MgSO_4$ was added (100mM stock) to bring the final concentration of Mg++ ions to 2mM, 3mM, 4mM, 5mM and 6mM respectively. The cycling parameters were 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 3 minutes, for 24 cycles. The reactions were analyzed by gel electrophoresis and the 3mM through 6mM Mg++ reactions were found to contain a DNA band of the desired size of 2.8kb. These reactions were pooled and the 2.8kb band was gel purified. The 2.8kb amplified MmeI gene fragment was digested with BamHI and PstI endonucleases (NEB) in the following reaction conditions:

15 µl 10X BamHI reaction buffer (NEB)
1.5 µl BSA (NEB)
50 µl MmeI gene 2.8 kb amplified DNA fragment
80 µl dH2O

5 µl BamHI endonuclease (100 units)
5 µl PstI endonuclease (100 units)

The reaction was mixed and incubated for 1 hour at 37°C. The small fragments cleaved off the ends of the 2.8kb DNA fragment were removed, along with the endonucleases, by purification on a Qiagen QiaPrep spin column according to the manufacturer's instructions.

The cleaved MmeI gene DNA fragment was ligated to the pRRS vector as follows: 10 µl of the digested, purified 2.8kb MmeI fragment was combined with 5 µl pRRS vector previously cleaved with BamHI and PstI and purified, 5 µl dH_2O, 20 µl 2X QuickLigase Buffer (NEB), the reaction was mixed, and 2 µl of QuickLigase was added. The reaction was incubated at room temperature for 5 minutes. 5 µl of the ligation reaction was transformed into 50 µl chemical competent *E. coli* ER2683 cells and the cells were plated on L-broth plates containing 100 µg/ml ampicillin and incubated at 37°C overnight. Approximately 200 transformants were obtained and 18 representatives were analyzed as follows: plasmid from each colony was isolated by miniprep procedures and digested with AlwNI and NdeI endonucleases to determine if they contained the correct size insert. 2 of the 18 transformants had the correct size insert of approximately 2800 bp. Both clones were tested to see if they produced MmeI endonuclease activity. The clones were grown overnight at 37°C in 500 mL L-broth containing 100 µg/ml ampicillin. The cells were harvested by centrifugation, suspended in 10 mL sonication buffer (20mM Tris-HCl, 1mM DTT, 0.1mM EDTA, pH7.5) and broken by sonication. The crude lysate was cleared by centrifugation and the supernatant was recovered. The lysate was assayed for endonuclease

activity by serial dilution of the lysate in 1X reaction buffer NEBuffer 1 (New England Biolabs) containing 20 µg/ml lambda DNA substrate and supplemented with SAM at 100 µM final concentration. The reactions were incubated for 1 hour at 37°C. The reaction products were analyzed by agarose gel electrophoresis on a 1% agarose gel in 1X TBE buffer. One of the two clones had MmeI endonuclease activity. This active clone was designated strain NEB1457 and was used for subsequent production of MmeI. The plasmid construct expressing MmeI activity in this clone was designated pTBMmeI.1.

EXAMPLE III

THE *Mme*I ENDONUCLEASE PROVIDES IN VIVO PROTECTION AGAINST MMEI CLEAVAGE

The plasmid pTBMmeI.1 was purified from NEB1457 using the Qiagen miniprep protocol. This plasmid has two MmeI sites in the vector backbone, and one site within the MmeI gene. The plasmid was digested with MmeI to test whether this DNA was resistant to MmeI endonuclease activity, which would indicate that the single MmeI gene was able to methylate DNA *in vivo* to protect the host DNA against its endonuclease activity. To test this the following were combined:

10 µl pTBMmeI.1 miniprep DNA
15 µl 10X NEBuffer 4
15 µl SAM (1mM stock solution)
110 µl dH2O
1 µl MmeI endonuclease (15 units)

The reaction was mixed and split in thirds. To one third was added 0.5 µl dH_2O, to the second was added 0.5 µl pRRS vector and to the third was added 0.5 µl PhiX174

DNA as a positive control. The pTBMmeI.1 was not cleaved by the MmeI endonuclease activity, while the Phix174 and pRRS DNAs in the same reaction were cleaved, indicating that the three MmeI sites in the pTBMmeI.1 DNA are resistant to MmeI endonuclease activity (Figure 4).

EXAMPLE IV

MmeI ENDONUCLEASE SENSITIVITY TO METHYLATION

The prior literature reports that MmeI endonuclease methylates just one strand of its recognition sequence, and that this hemi-methylation does not block subsequent cleavage of the DNA by the endonuclease (Tucholski, Gene 223 (1998) 293-302). To test this a set of four oligonucleotides were synthesized so that a DNA substrate could be formed that was either unmethylated (oligo 1 + oligo 2), methylated in the top strand only (oligo 3 + oligo 2), methylated in the bottom strand only (oligo 1 + oligo 4), or methylated on both strands (oligo 3 + oligo 4). The oligos synthesized were:

Oligo 1:

5'-FAM-GTTTGAAGACTCCGACGCGATGGCCAGCGATCGGCGCCTCAGCTTT TG-3' (SEQ ID NO:46)

Oligo 2:

5'-FAM-CAAAAGCTGAGGCGCCGATCGCTGGCCATCGCGTCGGAGTCTTCA AAC-3' (SEQ ID NO:47)

Oligo 3:

5'-FAM-GTTTGAAGACTCCG(6mA)CGCGATGGCCAGCGATCGGCGCCTCAGCTT TTG-3' (SEQ ID NO:48)

Oligo 4:

5'-FAM-CAAAAGCTGAGGCGCCGATCGCTGGCCATCGCGTCGG(6mA)GTCTTCA AAC-3' (SEQ ID NO:49)

(Other nucleotides outside the MmeI recognition sequence were also methylated for other studies, but since MmeI does not have any sequence specifity for these nucleotides this does affect MmeI activity and these other methylations are omitted here for clarity.) Duplex DNA was formed by mixing 100μl top strand oligo (14μM stock) with 100μl bottom strand oligo (14μM stock), heating to 85°C and cooling slowly to 30°C over a time of 20 minutes. MmeI was then used to cleave the oligo pairs in a 30 μl reaction of 1X NEBuffer4, 2.5 μM oligo, 100 μM SAM and 2.5 units MmeI. As a control, restriction endonuclease Hpy188I was also used to cleave the oligo DNA. The Hpy188I recognition sequence overlaps the first 5 nucleotides of the MmeI recognition sequence in this DNA, 5'-TCNGA-3' and is blocked by methylation at the adenine in either strand of the DNA. MmeI was found to cleave unmethylated DNA as expected. In contrast to previous teaching (Tucholski, Gene 223:293-302 (1998)) MmeI did not cleave the hemi-methylated DNA when the top strand only was methylated: 5'-TCCG(N6mA)C-3'. When the bottom strand only was methylated MmeI did cleave the DNA. When both strands were methylated MmeI did not cleave the DNA. (Figure 5) This finding is consistent with both the observed ability of the single MmeI enzyme to protect host DNA against cleavage in vivo and the observation that MmeI methylates only the top strand of its recognition sequence. We confirmed the report that MmeI enzyme methylates only the top strand of its recognition sequence by methylating the oligo

pairs above with tritium labeled H^3-SAM, washing away the unincorporated SAM and counting the radioactivity in the DNA. Both the unmethylated oligo DNA and the top unmethylated, bottom methylated DNAs had greater than 10-fold more counts than background, while the bottom unmethylated, top methylated DNA and the DNA with both strands methylated had counts near background (Figure 6). These findings indicate that MmeI is a novel type of restriction modification system which does not require a separate methyltransferase enzyme to modify the host DNA to provide protection against the activity of the endonuclease, as is the case for the type IIG (also called type IV) enzymes such as Eco57I.

EXAMPLE V

DNA SEQUENCING and ANALYSIS

DNA Sequencing: DNA sequencing was performed on double-stranded templates on an ABI 373 or ABI 377 automated sequencer. Amplified DNA fragments and individual clones were sequenced with primers synthesized as above or from universal primers located in the vector.

Computer analyses: Computer analyses of the DNA sequences obtained were performed with the Genetics Computer Group programs (Deverenx, et al., Nucleic Acids Res. 12:387-395 (1984)) and database similarity searches were performed via the internet at the National Center for Biotechnology Information site (http://www.ncbi.nlm.nih.gov/BLAST/) using the BLASTX and the BLASTP algorithms (Altschul, et al., J. Mol. Biol 215:403-410 (1990) and Gish, et al., Nature Genet. 3:266-722 (1993)).

WHAT IS CLAIMED IS:

1. Isolated DNA coding for the *Mme*I restriction enzyme, wherein the isolated DNA is obtainable from *Methylophilus methylotrophus*.

2. A recombinant DNA vector comprising a vector into which a DNA segment coding for the *Mme*I has been inserted.

3. Isolated DNA coding for the *Mme*I endonuclease and *Mme*I/methyltransferase, wherein the isolated DNA is obtainable from ATCC Accession No. PTA-4521.

4. Vectors that comprise the isolated DNA of claim 3.

5. A host cell transformed by the vector of claim 2 or 4.

6. A method of producing recombinant *Mme*I restriction endonuclease and *Mme*I methylase comprising culturing a host cell transformed with the vector of claims 2 or 4 under conditions suitable for expression of said endonuclease and methylase.

7. Isolated DNA coding for an *Mme*I-like restriction enzyme, wherein said isolated DNA hybridizes to at least one conserved motif of the nucleotide sequence coding for the *Mme*I restriction enzyme under predetermined conditions.

FIGURE 1: MmeI DNA CLEAVAGE

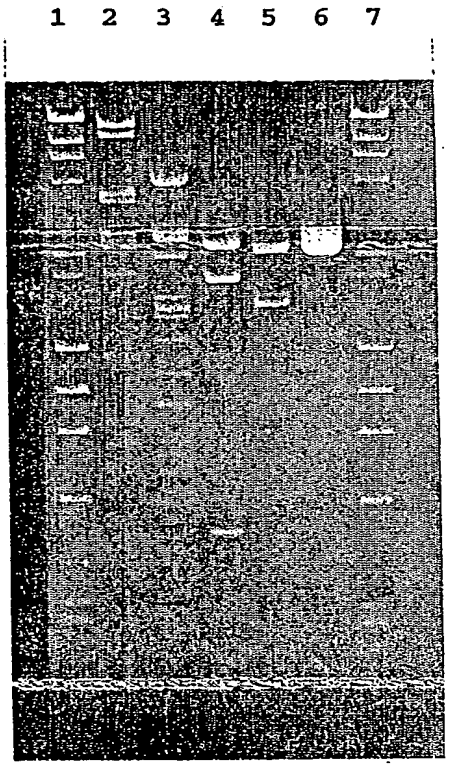

FIGURE 2: MmeI DNA SEQUENCE

```
   1  GAATTCCAGA TAGGTAGTCC TTTGGTACTT CCATCCCAAC CAGTGTCACG
  51  TTCCGCGCCA AACCAATCGG TTAAAGTGTA AGAAAGTCTT GCACTGAAGT
 101  AGCTGTAGGA CAAACCGAAG TTAACCTCTG TGGTATCCCA GCGACCACCT
 151  TTAGGTGTTT GACGGAAGCC TGCTGCGTCA CCTGCCAAGT TATATTTCTT
 201  CCATGAACCA CCTGGGTACA GGTAGCTGAT CAAACCAGCA GTCCAACCCA
 251  AGCCTTCAAT AGCAGGAATA GTTCCGTTAT ACCCACCATA AATATCAATT
 301  TCGGCAGTTG CATCAGGGAA GGTATTTGGT GTCACGTTTG AACCCCATGC
 351  ACCGACATAA AAGCCGCTGT CATGAGTAAT ATCAATACCG CCTTGAACGG
 401  CAGGTTTGTG CCAGTTTTGT GAAATACCAC GAGCATAGTA ATCTGAAACA
 451  AATCCAACGT TTGCAGTAGC AGCCCAGGCT GATTTTTCTT CTTTAGCCTC
 501  TTCAGCTGCG TATGAAACTT GGGCAAAAGA TAATGTGCTT AACACTGCTG
 551  TGAGCAATAT AGATTGACGC ATTATGAGTC CTCTCTCTGT GAAATCTTTG
 601  ATTAAGTTGT TGTAAACGAG AATGAAACAA CAACCACAAA GCAAAGCACG
 651  TGCCAAACTA TAAATAACAT TATAATCAAT TATTTAAAAT ATATTTATAA
 701  TCTAAAATAT TAAATTAATT ATTTAATAAA CTGTTTTTTA TTGATTTAAC
 751  TCTAAAACAT ATGGGTGCAA CCACCCTTTT TACTCACTGA TAATGCTAAN
 801  ATAGCCAACA AAGGAGCCTT CACCATGCTG ATTTCAAATG AAAAAATTCA
 851  GGAATTATCT TTAAAAATCA AACAACTAAT CGAATCAAGC CCCATTTCAG
 901  AGCTAAATAA CAACTTGCAT GCACTAATTC AGGGCGCACT CACCAAAATG
 951  GAACTTGTTT CGCGTGAAGA ATTCGATATC CAATCTGCAT TATTAGCGCG
1001  CACGCAAGAG CAATTAAAAC GTCTTGAAGA AAAAATCAGC CAGCTTGAAG
1051  AAGGGCAGGC ATCCAGAAAG TAAAAATTAA TTTACAATTG TTAGCATTCC
1101  ATTATTGAGG AGTGCGCTAT GAGTCTGGCG GTGTTATACA GTCGCGCGTT
```

1151 AAGCGGCATG GAGGCGCCAG AAGTGGTGGT AGAAGTCCAC TTGGCGAATG

1201 GACTACCCAG CTTTACCATT GTTGAAACAT ATTGAAACTT TAAGCCTTAG

1251 CATTTTTTCA AATATACAAA TGCCCCAAGC TGGTGCATTA AGAAGAATGT

1301 AACAACTCCC TGCAGACTAG GAATAACTTC ATGATTTAAC GAACATCCCT

1351 GAGTTTCAAA GTCGAATCTT CTCGTGTTGC AAATTTCTAC AGCTTCCTTT

1401 CTGACCCTCT TGCACCAAAT TGCACTATGG CGCTAATAAA TCTTCTGCTA

1451 TCCAATAATG TCCAACTAAC CCTTTATGGA CTCTTAAAAA AGATTTAATA

1501 AATGATTAAG ATGAATTCAA GGAATTTGAT GCCTGGAAAT ATGGCAAAAG

1551 CAAAAAGGCA GCCCAGTGCT GACTTTTTTG TTTTAACATT GGCCCATATA

1601 TCCAATTTCA AATAATTTAA AAATTATCGG GAGCTAATCT GTGGCTTTAA

1651 GCTGGAACGA GATAAGAAGA AAAGCTATTG AGTTTTCTAA AAGATGGGAA

1701 GACGCCTCAG ATGAAAACAG TCAAGCCAAA CCCTTTTTAA TAGATTTTTT

1751 CGAAGTTTTT GGAATAACTA ATAAGAGAGT TGCAACATTT GAGCATGCTG

1801 TCAAAAAGTT CGCCAAGGCC CATAAGGAAC AATCTCGAGG ATTCGTAGAT

1851 TTGTTTTGGC CTGGCATTCT TCTTATTGAA ATGAAAAGCA GAGGTAAAGA

1901 CCTCGACAAA GCGTATGACC AGGCACTTGA TTACTTTTCT GGCATTGCAG

1951 AAAGAGACTT ACCCAGATAC GTTTTAGTTT GCGACTTCCA GCGTTTCAGA

2001 TTAACAGACC TAATAACAAA AGAGTCAGTT GAATTTCTTT TAAAGGACTT

2051 ATACCAAAAT GTGAGGTCTT TTGGTTTTAT AGCTGGTTAT CAAACTCAAG

2101 TAATCAAGCC ACAAGACCCT ATTAATATTA AGGCGGCTGA ACGGATGGGT

2151 AAGCTTCATG ACACCCTGAA GTTGGTTGGA TATGAGGGAC ACGCTTTAGA

2201 ACTTTATCTA GTGCGTTTAC TTTTTTGCTT ATTCGCAGAA GACACAACTA

2251 TTTTTGAGAA AAGTTTATTC CAAGAATATA TCGAGACAAA GACGCTAGAG

2301 GACGGCAGTG ACCTTGCACA TCATATCAAT ACACTTTTTT ATGTTCTCAA

2351 TACCCCAGAA CAAAAAAGAT TAAAGAATCT AGACGAACAC CTTGCTGCAT

2401 TTCCATATAT CAATGGAAAA CTTTTCGAGG AGCCACTTCC GCCAGCTCAG

2451 TTTGATAAAG CAATGAGAGA GGCATTGCTT GACTTGTGCT CATTAGATTG

2501 GAGCAGGATT TCACCAGCAA TATTTGGAAG TTTATTCCAA AGCATTATGG

2551 ATGCTAAAAA GAGAAGAAAT CTTGGGGCAC ACTACACCAG CGAAGCAAAT

2601 ATTCTCAAGT TAATCAAGCC ATTGTTTCTT GACGAGCTCT GGGTAGAGTT

2651 CGAGAAAGTT AAAAATAATA AAAATAAATT ACTAGCGTTC CACAAAAAAC

2701 TAAGAGGACT TACATTTTTC GACCCTGCAT GCGGTTGCGG AAATTTTCTT

2751 GTAATCACAT ACCGAGAACT AAGACTTTTA GAAATTGAAG TGTTAAGAGG

2801 ATTGCATAGA GGTGGTCAAC AAGTTTTGGA TATTGAGCAT CTTATTCAGA

2851 TTAACGTAGA CCAGTTTTTT GGTATCGAAA TAGAGGAGTT TCCCGCACAG

2901 ATTGCTCAGG TTGCTCTCTG GCTTACAGAC CACCAAATGA ATATGAAAAT

2951 TTCAGATGAG TTTGGAAACT ACTTTGCCCG TATCCCACTA AAATCTACTC

3001 CTCACATTTT GAATGCTAAT GCTTTACAGA TTGATTGGAA CGATGTTTTA

3051 GAGGCTAAAA AATGTTGCTT CATATTAGGA AATCCTCCAT TTGTTGGTAA

3101 AAGTAAACAA ACACCGGGAC AAAAAGCGGA TTTACTATCT GTTTTTGGAA

3151 ATCTTAAATC CGCTTCAGAC TTAGACCTAG TTGCTGCTTG GTATCCCAAA

3201 GCAGCACATT ACATTCAAAC AAATGCAAAC ATACGCTGTG CATTTGTCTC

3251 AACGAATAGT ATTACTCAAG GTGAGCAAGT ATCGTTGCTT TGGCCGCTTC

3301 TGCTCTCATT AGGCATAAAA ATAAACTTTG CTCACAGAAC TTTCAGCTGG

3351 ACAAATGAGG CGTCAGGAGT AGCGGCGGTT CACTGCGTAA TTATCGGATT

3401 TGGGTTGAAG GATTCAGATG AAAAAATAAT CTATGAGTAT GAAAGTATTA

3451 ATGGAGAACC ATTAGCTATT AAGGCAAAAA ATATTAATCC ATATTTGAGA

3501 GACGGGGTGG ATGTGATTGC CTGCAAGCGT CAGCAGCCAA TCTCAAAATT

3551 ACCAAGCATG CGTTATGGCA ACAAACCAAC AGATGATGGA AATTTCCTAT

3601 TTACTGACGA AGAAAAAAAC CAATTTATTA CAAATGAGCC ATCTTCCGAA

3651 AAATACTTCA GACGGTTTGT GGGCGGGGAT GAGTTCATAA ACAATACAAG

3701 TCGATGGTGT TTATGGCTTG ACGGTGCTGA CATTTCAGAA ATACGAGCGA

3751 TGCCTTTGGT CTTGGCTAGG ATAAAAAAAG TCCAAGAATT CAGATTAAAA

3801 AGCTCGGCCA AACCAACTCG ACAAAGTGCT TCGACACCAA TGAAGTTCTT

3851 TTATATATCT CAGCCGGATA CGGACTATCT GTTGATACCT GAAACATCAT

3901 CTGAAAACAG ACAATTTATT CCAATTGGTT TTGTTGATAG AAATGTCATT

3951 TCAAGTAACG CAACGTATCA TATTCCTAGT GCTGAACCTT TGATATTTGG

4001 CCTGCTTTCA TCGACCATGC ACAACTGCTG GATGAGAAAT GTAGGAGGAA

4051 GGTTAGAAAG TCGTTATAGA TATTCTGCCA GCCTGGTTTA CAACACGTTT

4101 CCATGGATTC AACCCAACGA AAAACAATCG AAAGCGATAG AAGAAGCTGC

4151 ATTTGCGATT TTAAAAGCTA GAAGCAATTA TCCAAACGAA AGTTTAGCTG

4201 GTTTATACGA CCCAAAAACA ATGCCTAGTG AGCTTCTTAA AGCACATCAA

4251 AAACTTGATA AGGCTGTGGA TTCTGTCTAT GGATTTAAAG GACCAAACAC

4301 AGAAATTGCT CGAATAGCTT TTTTGTTTGA AACATACCAA AAGATGACTT

4351 CACTCTTACC ACCAGAAAAA GAAATTAAGA AATCTAAGGG CAAAAATTAA

4401 TTAATGTATT TAACATTAAA CCACCCTGAT TTATTTCGAA TAGTTCAAAT

4451 GCTTCCATGT GGACTAATCG CCTTCAATCA TATTAAAAAA CCGACGCTAG

4501 TAATAAAAAC TTCCAAAGAG GCCATATTAA CCGCCAAAAT TAATCGTGAA

4551 TTTAAAATAT ATCTTTATCA AACCACATCG GCTTGTGTTC TAGTAAGTGC

4601 ATTTTTTGAC GATTCTGATA GTCCACTATT CATTACAACA CCAATTGTTC

4651 GAGATGACCA ACACTCCTTA GACTTGTTAA GATTTTTAAT CAACAATGAT

4701 TTTACGATTT GCTTCTTTGA TGAACTGAAC CGAGAATTTC TTTCCGTTAA

4751 CGCAACTGGT AATTTAGTCT CTATCTTTGA GAGCATTCAC TTGATGCCAC

4801 TGCCGAGCCC AGAGGAAGCC CACAATGCAT TGAATGAAGC GGAATTTTGG

4851 TTCAGTTTAC GCTCAGCTGC TGATGATGAA TCATCTATCC AGGTTTCTTT

4901 ATTGGATAAT CTATTTCCTG ACGATTTTGT AATTTATGAC CTATCCTCAA

4951 ACAAAAACGA TATGACATCA TTGGTTAGAG AAACTAAACC AGGATACTAT

5001 CAGGAAGCAG ATATTGCAAA GTTACTAACA AGAGCTTTTA GTTTGGAAAG

5051 CATTTATCAG AATCCAGTGA AAACAAGCGA TTCAAAAGAG TTGGCAGACG

5101 TTGTGGTATT CGGCCAAAAG GAAATTTTAA TAATTCAAGC TAAAGATAGT

5151 GAAAACAATC AGAAACAAGT TTTAGAGGTT TCGTTAGACA AGAAATGCGC

5201 AAAGTCTTCA AAGAAACTTT CTGAAGCTTT GGCACAACTC ACCGACACTA

5251 TCTTAACAAT ATCCAATACA CCAATAGTTG ATGTTCGGGT TGGTAAGAAA

5301 AAATGCACTC TGAACTTTGA GGGAAAGCAG CTTATTGGTA TCGTCGTTGT

5351 TAAAGAGCTT TTTAATGATA TTTACGATAA ATACAGTCAA AAAGTTTTTG

5401 AGCATGTAGA GTTGTCTAAA GCACCCATTG TCTTCTTTGA CTATCCAGAA

5451 TTTGCAAGAA TGACATTTCA TTGTAATTCT GAGGAATTAT TACTTTATGC

5501 TTTGCATAGG ATATTTAGTT CTGCAATAGA AAATGGAATG TATAAACGAT

5551 TGAGATTTAC TCAACCTATC ATAACTGATG GTCATGACAG CTACTTCAGG

5601 ATACAAAACA GGCCCCATTC TGATGAGGCC TATTTAATTT GCACAGAGGA

5651 TGAAATGAAG CTCTCAAATA AGTTTAAAGA CTAAATTTAT ATTTTCCTCA

5701 GTATCTTAAA AACAATATTC ATTAAATTGG AAAGCCCGCA ATGATTGTTG

5751 CAGTATCAAT GCGGGCATCA GTATCCAGCT CTTGCAATAC ACGGAAGTAT

5801 CAAGAAGCGA ATCAGGATTC TAACCATACC TTTTTAATTG CAACAATCTA

5851 ATTTCCATAA CATGTGTAGC TACATCGAAA AAAAGACCTC GAAGAGGTTG

5901 CAAGAGCGTC CAGCTCGCGG CATCAAAAGA CCCTAGTCTT TTGACAAGGG

5951 GGAGCCAAAA AACTGAGGTG GAGGAGCTTG CCGACGAAGC CAGGAAGCCC

6001 CAGCGTCCGG

FIGURE 3: MmeI AMINO ACID SEQUENCE

```
   1  MALSWNEIRR KAIEFSKRWE DASDENSQAK PFLIDFFEVF GITNKRVATF
  51  EHAVKKFAKA HKEQSRGFVD LFWPGILLIE MKSRGKDLDK AYDQALDYFS
 101  GIAERDLPRY VLVCDFQRFR LTDLITKESV EFLLKDLYQN VRSFGFIAGY
 151  QTQVIKPQDP INIKAAERMG KLHDTLKLVG YEGHALELYL VRLLFCLFAE
 201  DTTIFEKSLF QEYIETKTLE DGSDLAHHIN TLFYVLNTPE QKRLKNLDEH
 251  LAAFPYINGK LFEEPLPPAQ FDKAMREALL DLCSLDWSRI SPAIFGSLFQ
 301  SIMDAKKRRN LGAHYTSEAN ILKLIKPLFL DELWVEFEKV KNNKNKLLAF
 351  HKKLRGLTFF DPACGCGNFL VITYRELRLL EIEVLRGLHR GGQQVLDIEH
 401  LIQINVDQFF GIEIEEFPAQ IAQVALWLTD HQMNMKISDE FGNYFARIPL
 451  KSTPHILNAN ALQIDWNDVL EAKKCCFILG NPPFVGKSKQ TPGQKADLLS
 501  VFGNLKSASD LDLVAAWYPK AAHYIQTNAN IRCAFVSTNS ITQGEQVSLL
 551  WPLLLSLGIK INFAHRTFSW TNEASGVAAV HCVIIGFGLK DSDEKIIYEY
 601  ESINGEPLAI KAKNINPYLR DGVDVIACKR QQPISKLPSM RYGNKPTDDG
 651  NFLFTDEEKN QFITNEPSSE KYFRRFVGGD EFINNTSRWC LWLDGADISE
 701  IRAMPLVLAR IKKVQEFRLK SSAKPTRQSA STPMKFFYIS QPDTDYLLIP
 751  ETSSENRQFI PIGFVDRNVI SSNATYHIPS AEPLIFGLLS STMHNCWMRN
 801  VGGRLESRYR YSASLVYNTF PWIQPNEKQS KAIEEAAFAI LKARSNYPNE
 851  SLAGLYDPKT MPSELLKAHQ KLDKAVDSVY GFKGPNTEIA RIAFLFETYQ
 901  KMTSLLPPEK EIKKSKGKN*
```

FIGURE 4: pTBMmeI.1 IS RESISTANT TO MmeI CLEAVAGE

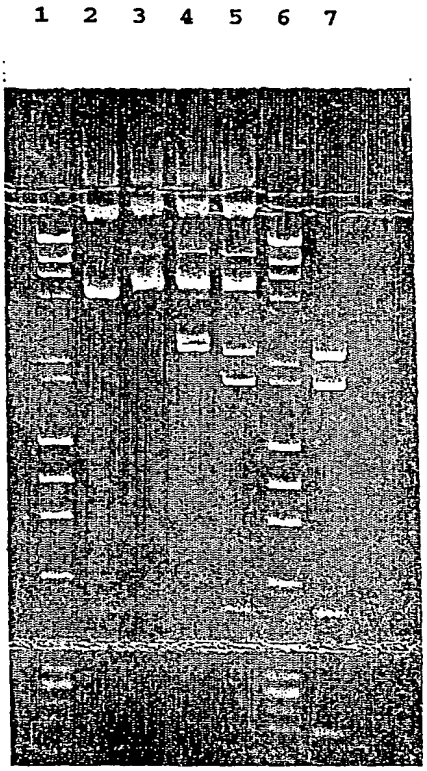

FIGURE 5: MmeI CLEAVAGE OF HEMI-METHYLATED SUBSTRATES

1 2 3 4 5 6 7 8 9 10 11 12 13 14

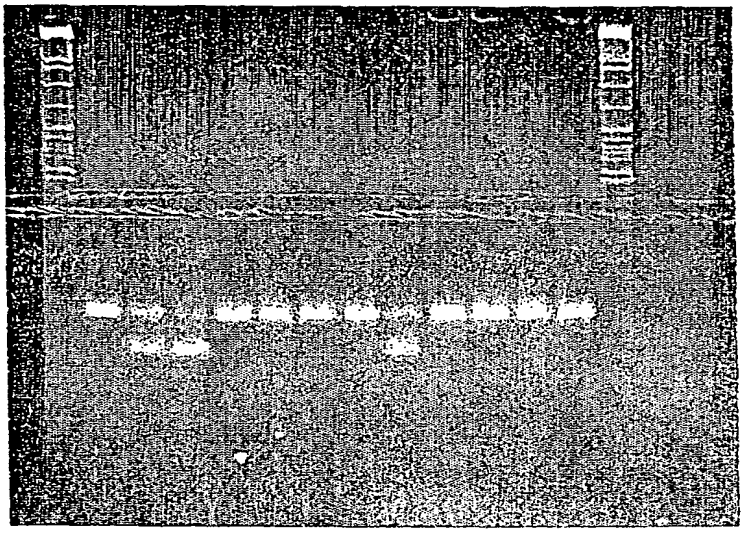

FIGURE 6: METHYLATION INCORPORATION BY MmeI ENDONUCLEASE

Top Strand: / Bottom strand:	^{3}H-COUNTS
5'-TCCG**A**C-3' / 5'-GTCGG**A**-3'	
unmethylated: / unmethylated	19,972
unmethylated: / methylated:	14,447
methylated: / unmethylated:	1,266
methylated: / methylated:	917

'**A**' indicates position of N6-methyl adenine in the DNA substrate

MmeI Figure 7: Multiple Sequence Alignment of MmeI and Homologous Polypeptides

PileUp of: @mme.list2

Symbol comparison table: GenRunData:blosum62.cmp CompCheck: 1102

GapWeight: 6
GapLengthWeight: 1

Name: mmelfeLORF3P = gi|28373198|ref|NP_783835.1| (SEQ ID NO:3)
Name: mmeLre121P = gi|23451826|gb|AAN32874.1|AF461726_1 (SEQ ID NO:4)
Name: mme = MmeI amino acid sequence (SEQ ID NO:5)
Name: mmeNMA1791 = gi|15794682|ref|NP_284504.1| (SEQ ID NO:6)
Name: mmeBSU0677 = gi|16077744|ref|NP_388558.1| (SEQ ID NO:7)
Name: mmegcry = gi|9945797|gb|AAG03371.1| (SEQ ID NO:8)
Name: mmePflQ8 = gi|23451826|gb|AAN32874.1|AF461726_1 (SEQ ID NO:9)
Name: saro3834 = gi|23110638|gb|ZP_00096791.1| (SEQ ID NO:10)
Name: mmeMSI135 = gi|20803963|emb|CAD31540.1| (SEQ ID NO:11)
Name: mmeCC0826 = gi|16125079|ref|NP_419643.1| (SEQ ID NO:12)
Name: mmeDR0119.1 = gi|15807788|ref|NP_285443.1| (SEQ ID NO:13)
Name: mmeDR2267 = gi|15807258|ref|NP_295988.1| (SEQ ID NO:14)

```
                1                                                  50
mmelfeLORF3P  ~~~~~~~~~~ ~~~~~~~MPT RQQAAREFVK TWS.SDKKGR EDADRQTFWN
  mmeLre121P  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
         mme  ~~~~~~~~~~ ~~~MALSWNE IRRKAIEFSK RWE.DASD.. ENSQAKPFLI
  mmeNMA1791  ~~~~~~~~~~ ~~~~MKTLLQ LQTAAQNFAA YYK.DQTD.. ERREKDTF*N
  mmeBSU0677  ~~~~~~~~~~ ~~~~~MALID LEDKIAEIVN R.E.DHSD.. .......FLY
     mmegcry  MVMAPTTVFD RATIRHNLTE FKLRWLDRIK QWEAENRPAT ESSHDQQFWG
    mmePflQ8  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     saro3834 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
    mmeMSI135 ~~~~~~~~~~ ~~MSLGAAGL TPITPAAFIK KWRKSELG.. ERQAAQEHFL
    mmeCC0826 ~~~~~~~~~~ ~~~~~~~~~~ ~~MTPAQFVK KWSDSQLR.. ERQASQEHFL
  mmeDR0119.1 ~~~~~~~~~~ ~~~~~~~~~~ ~~MHPQEFAD TWSRRALKAT ERDSYVQHWL
    mmeDR2267 ~~~~~~~~~~ ~~~~~MPQTE TAQRMEDFVA YW..RTLKGD EKGESQV.FL

                51                                                 100
mmelfeLORF3P  DLLQRVYGID N.YYDYITYE KDVQVKADGK VTTRRIDGYI P.STKIMVEM
  mmeLre121P  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
         mme  DFFE.VFGIT N.KR...VAT FEHAVKKFAK AHKEQSRGFV DL...FWPGI
  mmeNMA1791  EFFA.IFGID R.KN...VAH FEYPVKD..P ADNTQ...FV DI...FWEGI
  mmeBSU0677  ELLG.VYDVP R.AT...ITR ....LKK.GN QNLTKRVGEV HLKNKVW...
     mmegcry  DLLDC.FGV. N.ARDLYLY. .....QRSAK RASTGRTGKI DM...FMPGK
    mmePflQ8  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
     saro3834 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~M
    mmeMSI135 D.ICSLVGHP SP.SDEDPTG AFFAFEKGAN KLG.GGKGFA D.VWK..KGH
    mmeCC0826 D.LCRMLEVP TP.AEDDPLG ERYCFERGAA KTG.GGDGWA D.VWR..KGC
  mmeDR0119.1 D.LCQLLHHE APGADPD... ..YKFERRVT KVGTKDKGFA D.VFK..KAH
    mmeDR2267 DRLFQAFGH. ...AGYKEAG AE..LEYRVA KQG.GGKKFA DLLWR..PRV

                101                                                150
mmelfeLORF3P  KGKNIKDLSK PITQSGGD.. .......ELT PFEQAKRYAN FLPN...SEQ
  mmeLre121P  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
         mme  LLIEMKSRGK DL.....D.. .......KAY D..QALDYFS GIAERD...L
  mmeNMA1791  FLAEHKSANK NL.....T.. .......KAK E..QAERYLQ EIGRTKPSAL
  mmeBSU0677  .FKERK.KGK LF.....D.. .......ALI DIEQQVEYL. .......SAK
```

```
      mmegcry  VIGEAKSLGV PLDDA..... .......... .YAQALDYLL G.GTIANSHM
      mmePflQ8 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      saro3834 NPVEIEEAVS DLARAPYDAS EFPFQFLAAF GNKQTTLQRL RAGNSNQSDL
      mmeMSI135 FAWEYKRKKG NLDEA..... .......... .LLQLMRYAP AL........
      mmeCC0826 FGWEYKGKHK NLDAA..... .......... .LRQLQAYAL DL........
      mmeDR0119.1 FITEYKRPGS DLGAA..... .......... .LQQATLYSR DL........
      mmeDR2267 LI.EMKKRGE KLANH..... .......... .YQQAFDYWL KL........

                151                                                 200
      mmelfeLORF3P PR........ WILVSNFNEI DIHDM..E.. .RPLDEPKVI KL........
      mmeLre121P ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      mme      PR........ YVLVCDFQRF RLTDLITK.. .ESVE....F LL........
      mmeNMA1791 PE........ YYAVSDFAHF HLYRRVPE.. .EGAENQWQF PL........
      mmeBSU0677 PR........ YLLVTDYDGV LAKDTKTL.. .EALDVKF.. ..........
      mmegcry  PA........ YVVCSNFETL RVTRLNRTYV GDSADWDITF PL........
      mmePflQ8 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      saro3834 PGAVLQRNHI HIATCDAGNV DRTLAALRKS PKTASQKARF ILATDGVAFQ
      mmeMSI135 ....L..SPP LHIVCDIERL RIHTAWTNTV PSTY..VITL DDLAE.....
      mmeCC0826 ....Q..NPP YLVVSDMERI IVHTNWTNTI SRKI..EFTL DDLHE.....
      mmeDR0119.1 ....G..NPP LLLTSDFQRI EINTAFTGTS PKSY..LITL DDIAENRVVG
      mmeDR2267 ....VPDRPR YAVLCNFDEL .....WVYDF NQQ.....L DEPMDRLRI.

                201                                                 250
      mmelfeLORF3P .EDLPKKVKS L........E F.....MVDA NQQQVIDEKQ LSVDAGNLVA
      mmeLre121P ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      mme      .KDLYQNV.R S......... FGFIA...GY QTQVIKPQDP INIKAAERMG
      mmeNMA1791 .EELPEYITR G........V FDFMF...GI EAKVRQIQEE ANIQAAATIG
      mmeBSU0677 .EELPQY... .......... FDFFLAWKGI EKVEFEKENP ADIKAAERFA
      mmegcry  .AEIDEHIEQ L........A F.....LADY ETSAYREEEK ASLEASRLMV
      mmePflQ8 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      saro3834 AEDMASGETV ACNYAAFPDK FAFFLPLAGI TTVQQIRESS FDIKATGRLN
      mmeMSI135 ....PSAREM LHNVFFSPEK .........L RPTR..TRAA VTKEAADKFS
      mmeCC0826 ....PEKLAM LRQVFDGSDS .........L KPKI..SPQE LTAKVAQRFG
      mmeDR0119.1 GNDVP.ALQI LHSALHQPYD .........L DPRL..FRER ITTDATRQVG
      mmeDR2267 .EELPERYTV LNFMFEQ..E .........R APLFGNNRVD VTREAADSVA

                251                                                 300
      mmelfeLORF3P KIYNELTNAY AAGRGIDVN. ...EPRIQRS LN..MLIVRL VFLLYADDSN
      mmeLre121P ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      mme      KLHDTL.K.. ......LVG. ....YE.GHA LE..LYLVRL LFCLFAEDTT
      mmeNMA1791 RLHDAL.K.. ......EEG. ...IYE.EHE LR..LFITRL LFLFFADDSA
      mmeBSU0677 RIYDVLRK.. ......ENN. ...IIETNRG LD..LFLIRL LFCFFAEDTD
      mmegcry  ELFRAMNGDD VDEAVGDDAP TTPEEEDERV MRTSIYLTRI LFLLFGDDAG
      mmePflQ8 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      saro3834 KLYVELLKDN PDWA...... .......SRS EDMNHFMARL IFCFFAEDTD
      mmeMSI135 AIALRVQGR. G.TPD..... .......... .EIAHFVNQL VFCFFAQSVS
      mmeCC0826 DLGRRLQER. GHHPR..... .......... .DVAHFLNRV VFCMFAEDAK
      mmeDR0119.1 LVARRLGERE GRT..... .......... .RAAHMMMRV VFALFAEDTG
      mmeDR2267 KVLNSVIAR. GEDRA..... .......... .RAQRFLLQC VMAMFAEDFE

                301                                                 350
      mmelfeLORF3P LFGKEDIFQA FIER...REP RDIRRDLSEL FKVLDQP.EE QRDPYLDDEF
      mmeLre121P ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
      mme      IFEKS.LFQE YIETKTLEDG SDLAHHINTL FYVLNTP.EQ KRLKNLDEHL
      mmeNMA1791 VFRRNYLFQD FLE..NCKEA DTLGDKLNQL FEFLNTP.DQ KRSKTQSEKF
      mmeBSU0677 IFKRNS.FTN LIKTLTEEDG SNLNKLFADL FIVL....DK NERDDVPSYL
      mmegcry  LWDTPHLFAD FVRNETTPE. .SLGPQLNEL FSVLNTA.PE KRPKRLPSTL
      mmePflQ8 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
```

```
saro3834        IFVGEGLFSR TVETMSARDA SDTHMVIAEI FRAMDTRLAD RAAAGIKSWA
mmeMSI135       LLPD.GLFTK LLK.RSARAP ERAMSYLDKL FEAME..... RGGEF...DL
mmeCC0826       LLPE.GLFTR LTRSMQMRPP AEAAPQFDAL FAMMR..... AGGMF...GA
mmeDR0119.1     MLER.GIVTR LLE.RARAPP GEDQLYFQDL FGAMK..... GGGEF...WG
mmeDR2267       LIPR.GFFTE LADD.ARAGR GSSFDLFGGL FRQMNTSERA RGGRF.....

                351                                                 400
mmelfeLORF3P    NQFAYVNGGM FSDENVIIPQ FTDELKRLIV EDAGRGFDWS GISPTIFGAV
mmeLre121P      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
mme             AAFPYINGKL FEEPLPPA.Q FDKAMREALL .DLCS.LDWS RISPAIFGSL
mmeNMA1791      KGFEYVNGGL FKERLRTF.D FTAKQHRALI .DCGN.FDWR NISPEIFGTL
mmeBSU0677      KEFPYVNGQL FTEPHTEL.E FSAKSRKLII .ECGELLNWA KINPDIFGSM
mmegcry         AKFPYVNGAL FAEPLAS.EY FDYQMREALL AAC..DFDWS TIDVSVFGSL
mmePflQ8        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
saro3834        DVFPYVNGQL FSGS.TECPR FSKIARSYLL H..IGSLDWQ KINPDIFGSM
mmeMSI135       TDITWFNGGL FDGR..RALR LDDGDIGLL. .VAADSLDWG LIDPTIFGTL
mmeCC0826       DIVHWFNGGL FDEK..PALP LERADIKLIH DTAAEH.DWS DLDPSVFGNM
mmeDR0119.1     TDIRHFNGGL FDSE..DALA LTSEDAAAL. .IIAAKLDWS EVEPSIFGTL
mmeDR2267       APIPYFNGGL FRAV..DPIE LNRDELYLLH KAALEN.NWA RIQPQIFGVL
                401                                                 450
mmelfeLORF3P    FESTLN.PET RRSGGMHYTS IENIHKVIDP LFLNDLHDEF D.........
mmeLre121P      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
mme             FQSIMD.AKK RRNLGAHYTS EANILKLIKP LFLDELWVEF E.........
mmeNMA1791      FQSVMD.AQE RREAGAHYTE AANIDKVING LFLENLRAEF E.........
mmeBSU0677      IQAVAS.EES RSYLGMHYTS VPNIMKVIKP LFLDKLNQSF ..........
mmegcry         FQLVKS.KEA RRSDGEHYTS KANIMKTIGP LFLDELRAEA D.........
mmePflQ8        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
saro3834        IQAVAD.DEE RGALGMHYTS VPNILKVLNP LFLDDLRAKL E.........
mmeMSI135       FERFLD.PEK RAQIGAHYTD PEKIMRLVDP VILRPLRQEW EQARREIVEL
mmeCC0826       FEEALKATRE RAALGAHYTD REKILKIIDP VITWPLMAQW ETALAEIRAA
mmeDR0119.1     FENSLDV.DT RSRRGAHYTS VNDIERIVDR VVMEPLWAEW D.........
mmeDR2267       FQSSMDKKEQ HAK.GAHYTS EADIMRVVLP TIVTPFQRQI EAATTQ....

                451                                                 500
mmelfeLORF3P    .......... .......... .......... .......... ....KIQNMG
mmeLre121P      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
mme             .......... .......... .......... .......... ....KVK...
mmeNMA1791      .......... .......... .......... .......... ....AVK..A
mmeBSU0677      .......... .......... .......... .......... .........L
mmegcry         .......... .......... .......... .......... ....KL..VS
mmePflQ8        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
saro3834        .......... .......... .......... .......... ....E.....
mmeMSI135       LNGN...... .......... ..RKPPMRR. ..QQSRR... ...MKREEAA
mmeCC0826       LDARAAAEAE RKAVLEAAAE AMRADPVKAK AGEAARRKTL TAIAKRSDAA
mmeDR0119.1     .......... .......... .......... ....ALRLSL PELKK.....
mmeDR2267       .......... .......... .......... .......... ..........

                501                                                 550
mmelfeLORF3P    NRRQRVTRAK AFRDKLGKLK FFDPACGSGN FLTETYLSLR KMENECLRII
mmeLre121P      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
mme             ...NNKNKLL AFHKKLRGLT FFDPACGCGN FLVITYRELR LLEIEVL.RG
mmeNMA1791      LKRDKAKKLA AFYQKIQNLQ FLDPACGCGN FLIVAYDRIR ALEDDIIAEA
mmeBSU0677      DAYDDYTKLE NLLTRIGKIK FFDPACGSGN FLIITYKELR RMEINIIKRL
mmegcry         SPSTSVAALE RFRDSLSELV FADMACGSGN FLLLAYRELR RIETDIIVAI
mmePflQ8        ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
saro3834        .AGDNSRKLL NLRNRMAKIR VFDPACGSGN FLVIAYKQMR ELEAEI....
mmeMSI135       .AEVR.SR.. .FTERLRKLR ILDPACGSGN FLYLALQGVK DIEHRANLDC
mmeCC0826       LGQAK.DRLE AFLSKLAAFR VLDPACGSGN FLYVALHALK DIERRALVDA
```

```
                751                                                      800
mmelfeLORF3P  VLFNDFHIFI NYGYRSFEWN NEAANKAKVD VVIVGFSTK. .EDKNPTIYD
mmeLre121P    ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~VNKDKILYN
mme           LLLS.LGIKI NFAHRTFSWT NEASGVAAVH CVIIGFGLKD ..SDEKIIYE
mmeNMA1791    SLLN.QGIEI HFAHRTFQWT SQAAGKAAVH CIIVGFRQKP PMPSEKTLYD
mmeBSU0677    ELFK.FGIQI NFAYKSFKWA NNAKNNAAVI VVIVGFG... PLDTKVNKYL
mmegcry       PIFKA.GWRI RFAHRTFAWD SEAPGKAAVH CVIVGFDKES ..QPRPRLWD
mmePflQ8      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
saro3834      RILKSAN..V KFAYRPFRWS NSAANNAGVY CTIIGLTGSE VSNKK.....
mmeMSI135     RIIA.ES.RL FEAWSDEPWV VDG...AAVR VSLICFGHG. .EDPLCL...
mmeCC0826     PIAD.AG.AL MEAWADEPWA LEG...AAVR VSMFGFGDG. .FAERRL...
mmeDR0119.1   RIKA.TG.DL FMAWPDEPWQ QNG...AAVR VSLFGFDNG. .TETLRT...
mmeDR2267     YVVQHGG.TI TDAVGTQVWS GD....AAVH VSIVNWVKGP AEGPKHLAWQ

                801                                                      850
mmelfeLORF3P  EQKIIS.... ..A.KHINQY MYDSDNIFID TTRKY.IEA. MPKMKTGNRP
mmeLre121P    SSN*ISH... ..C.KNINGY LFDGNNIFV. TNRPAPLSN. VPRMHNGCKL
mme           YESINGEPLA IKA.KNINPY LRDGVDVIA. CKRQQPISK. LPSMRYGNKP
mmeNMA1791    YPDIKGEPEK HAV.ANINPY LIDAPDLII. AKRSRPIHC. EPDMVNGSKP
mmeBSU0677    FVD...ETKK L.V.SNISPY LTDGENILV. SSRTKPISD. LPKLHFGNMP
mmegcry       YPDVKGEPVS VEVGQSINAY LVDGPNVLVD KSR.HPISSE ISPATFGNMA
mmePflQ8      ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~P.ADE RSQMDAGGKP
saro3834      ...LFGEGSV VEC.SSIAPY LVPGPEI.IC APRQSSIS.G FARMVMGSNP
mmeMSI135     ..DGRT.... ...VAQINAD LTAGVTDLTK ARRL.S..EN QNVAFMGDTK
mmeCC0826     ..EGRK.... ...AEHLHSD FRGASTDVTK ALRL.K..EN ASIAFMGDTK
mmeDR0119.1   LNDGH..... ...VGVINAD LNAG.TDVKQ AQKL.P..EN AGVSFIGTQK
mmeDR2267     VGDHRTSPWQ STELPVINSA LSAG.TDVTQ AQKL.RVNMN SGACYQGQTH

                851                                                      900
mmelfeLORF3P  ADGGALILSP KEAKEL.VNE EPQSKQ...F IKKLTGSKEF IT.GKYRYCL
mmeLre121P    LDGGFYTLTS QERKEA.ISK DPYADK...F IRPYLGAKNF IH.GTARYCI
mme           TDDGNFLFTD EEKNQF.ITN EPSSEK...Y FRRFVGGDEF IN.NTSRWCL
mmeNMA1791    TEGGNLILST AEKDAL.IAA EPLAEQ...Y IRPFIGADEF LN.GKTRWCL
mmeBSU0677    NDGGGLLFTI TEYTDA.INK YPELVP...Y FKKFIGSVEF IN.GGLRYCL
mmegcry       RDGGNLLVEV DEYDEV.MS. DPVAAK...Y VRPFRGSREL MN.GLDRWCL
mmePflQ8      VEGGNLLFAE EEKQRL.VEG NVDVVK...F LKRVYGASEY IR.GEVRFCL
saro3834      VDGKRLIFEQ DEKESV.VAA DPRSER...F FKRYGGTQEL VN.GVDRWCL
mmeMSI135     G.GAFDVPGS LARAWLSMPM NPNGRPNSDV LRPWRNGMDV ARRG...RDM
mmeCC0826     G.GAFDVSGE IAREWLRLPL NPNGRPNSDV LKPWRNAMDM TRRS...SDK
mmeDR0119.1   G.GAFDIPGD LARSWLSVP. NPDRVSNADV LKPWVNGMDL TRRP...SGR
mmeDR2267     GHKGFLLDGL EAGQMLSAE. ....RKNAEV IFPYLTGDEL LRTSPPHPTR

                901                                                      950
mmelfeLORF3P  WLVNVTP... .KQLR..SMP L..VLKRVE. QCKENR.LSG APDRQKLAAT
mmeLre121P    WLKDANP... .KDIH..QSP F..ILDRIN. KVAEFRSQQK SKDTQKYAKR
mme           WLDG.ADIS. ..EIR..AMP L..VLARIK. KVQEFRLKSS AKPTRQSAST
mmeNMA1791    WFHGVSDVKR NHDLK..QMP Q..VQARIQ. AVKTMREASS DKQTQKDAAT
mmeBSU0677    WLNE....AK YEKIK..SNP L..IQERIS. ISKNHREKST DKGTNKLALT
mmegcry       WLVDVAP... .SDIA..QSP V..LKKRLE. AVKSFRADSK AASTRKMAET
mmePflQ8      WISD....SQ EQEAK..SNS D..INCKLN. AVAAFRLKSP KAATKKGAAW
saro3834      WIND....DQ VDDAK..AIA E..IAKVLE. SCRSYR.QGA GRDAQKAANR
mmeMSI135     WIVDFGWEMS EQEAALYEAP FQHIREHVFP .........ER SKN.......
mmeCC0826     WIIDFGWTMS EADAALFETP FRHVLLHVKP .........ER DRN.......
mmeDR0119.1   WIIDFA.QMD EGEARQYLQP MAYVEQKIRP .........ER ATN.......
mmeDR2267     YVIDF.QPRD VFGARAYKLP FARIEREVLP TRQAAAAEEE ARNAEVLAAN

                951                                                     1000
```

```
mmelfeLORF3P  P.........  .........HL  ......FREQ  MNPDNYMIVP  LVTGCRRKYV
  mmeLre121P  P.........  .........ML  PTRLAYYSHD  EHTD.MLIVP  ATSSQRREYL
         mme  P.........  .........M.  ..KFFYISQP  .DTDY.LLIP  ETSSENRQFI
  mmeNMA1791  P.........  .........W.  ..LFQKIRQP  SDGNY.LIIP  SVSSESRRFI
  mmeBSU0677  P.........  .........W.  ..KFRDTHE.  .TTNYSIVVP  SVSSENRFYI
     mmegcry  P.........  .........HL  FGQ....RSQ  PDTDY.LCLP  KVVSERRSYF
   mmePflQ8   P.........  .........H.  ..KFEEVKQI  GN.EVVTIVP  KVSSESREYL
     saro3834 P.........  .........HS  FC.YRTFQE.  ...NIGIHVG  LTIGNGLSHV
   mmeMSI135  ....RRDAYR  ...ERWWRHV  EPRPAFHASL  QGHSRYMATP  RV.AKHRTFV
   mmeCC0826  ....NREMYR  ...LNWWKHV  EPRQGLMKRV  PALSRLLVTP  EV.SKHRLFI
 mmeDR0119.1  ....SDRPSR  ...ERWWLHQ  RSRPELREAT  IELDRFIGIP  RV.AKHLLPV
  mmeDR2267   PKAKTNKHHR  NFLNQWWALS  YGRSEMIEKI  SSLSRYIVCS  RV.TKRQVFE

                1001                                                1050
mmelfeLORF3P  PFGYLG.NDI  IPTNLATIIP  EADHYAFGVL  ESIVHMAWMR  VVA...GRKG
  mmeLre121P  PIGYVSEKNI  VSYSL.MLIP  NASNFNFGIL  ESKVHYIWLK  NFC...GRLK
         mme  PIGFVDRNVI  SS.NATYHIP  SAEPLIFGLL  SSTMHNCWMR  NVG...GRLE
  mmeNMA1791  PIGYLSFETV  VS.NLAFILP  NATLYHFGIL  SSTMHNAFMR  TVA...GRLK
  mmeBSU0677  PMGLAGADTI  LS.NLIYVIY  DAEIYLLGIL  MSRMHMTWVK  AVA...GRLK
     mmegcry  TVQRYPSNVI  AS.DLVFHAQ  DPDGLMFALA  SSSMFITWQK  SIG...GRLK
   mmePflQ8   PVGLLPRGSI  VT.DLAFALY  DAPLWNMALI  ASRLHLVWIG  ~~~~~~~~~~
     saro3834 PAD.LKSSGF  VSSHTAYMIY  GWHPVEFALL  NSRLMLVWTE  TVG...GRLG
   mmeMSI135  WLD....QAI  VPDSRIFAFS  RSDDVFFGIL  HSRFHEAWSF  GTCSWHGV.G
   mmeCC0826  WLD....ARV  LPDHKLQVVT  LDDDCSFGVL  HSRFHEVWAL  AAGSWHGS.G
 mmeDR0119.1  WLP....EGT  LPDSQVVVIA  RDDDFIFGVL  ASTIHRSWAR  MQGTYMGV.G
  mmeDR2267   FLD....NGI  RPSDGLQIFA  FEDDYSFGVI  QSSVHWQWLI  ARG...GTLT

                1051                                                1100
mmelfeLORF3P  TSYRYSKNLV  YTNFPWPV.V  DINQKEK...  .......ITI  TAQDILNARN
  mmeLre121P  SDYRYSNTII  YNNFPWPT.V  GDKQEQN...  .......ISE  TAQGILNTRK
         mme  SRYRYSASLV  YNTFPWIQ..  ....PNEKQ.  ....SKAIEE  AAFAILKARS
  mmeNMA1791  SDYRYSNTVV  YNNFPFPE..  SCRLPSENDR  PDPLRAAVEA  AAQTVLDARG
  mmeBSU0677  TDYRYSAGLC  YNTFPIPE.L  STRRKNE...  .......IEE  AILEILDLR.
     mmegcry  SDLRFANTLT  WNTFPVPE.L  DEKTRQR...  .......IIK  AGKKVLDARA
   mmePflQ8   ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~
     saro3834 NGMRFSNTIV  YNTFPVPS.L  TDQNK.....  .....ADLTR  CAEDILLARE
   mmeMSI135  NDPTYNSAGV  FETFPFPEGL  TPDIPAVRYE  KDSRAIAISK  AAKRLDDIRN
   mmeCC0826  NDPRYTISTT  FETFPFPEGL  TPNIAAVDYE  GDPRAQAIAA  AAAELNRLRE
 mmeDR0119.1  NDLRYTPSTC  FETFPVP...  APT.......  .DEQRAEIEK  WAKYIVQLRE
  mmeDR2267   ARLMYTSDTV  FDTFPWP*D.  .PTLAQVR..  ......AVAA  AAVKLRELRN

                1101                                                1150
mmelfeLORF3P  ..........  ..........  ...LY.....  ....PDSSLA  DLYDPLTMPI
  mmeLre121P  ..........  ..........  ...LY.....  ....PDSSLA  DLYDPLTMPV
         mme  ..........  ..........  ...NY.....  ....PNESLA  GLYDPKTMPS
  mmeNMA1791  ..........  ..........  ...QYRREAQ  EAGLPEPTLA  ELYAPDAGYT
  mmeBSU0677  ..........  ..........  ........EE  QGG....TLA  ELYNPSTMPI
     mmegcry  ..........  ..........  ...LH.....  ....PERSLA  EHYNPLAMAP
   mmePflQ8   ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~  ~~~~~~~~~~
     saro3834 ..........  ..........  ...SHFP...  ......ATIA  DLYDPETMPE
   mmeMSI135  AWLNPSDLVQ  IKPEVVPGYP  DRILPKDIAS  DAILRDRTLT  NLYNR.....
   mmeCC0826  AWLNPPDLVR  IEPEVVPGYP  DRVLPVSPEA  GAELKKRTLT  NLYNQ.....
 mmeDR0119.1  HLLN......  ..........  .........Q  DA...KGTLT  GIYNQLEKLR
  mmeDR2267   KVMREQ....  .......GW.  ..........  .......SLR  DLYRTLDMPG

                1151                                                1200
mmelfeLORF3P  E.........  .LRKAHEAND  KAVLKAYGLK  P.SATEPEI.  .VQHLFKMYE
  mmeLre121P  E.........  .LRKAHEAND  KAVLKAYGLS  P.KATEQEI.  .VEHLFKMYE
```

```
                mme  E......... .LLKAHQKLD KAVDSVYGFK .GPNT..EIA RIAFLFETYQ
          mmeNMA1791  A......... .LDKAHATLD KAVDKAYGYK TGKNTDDEAE RVAFLFELYR
          mmeBSU0677  E......... .LKVAHEKLD GIVERAYRQK QFES..DE.E RLEVLLKLYQ
             mmegcry  E......... .LIKAHDALD REVDKAFGAP RKLTTVRQ.. RQELLFANYE
            mmePflQ8  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
             saro3834 S......... .LRAAHDRND EVLERIY... IGRRFRNDTE RLEKLFELYT
          mmeMSI135   .......RP.Q WLVDAHSDLD AAVAGAYGWP ADIS...EDE ALANLLELNL
          mmeCC0826   .......RP.A WLDMAHQRLD AAVAAAYGWP DGLT...DDE ILERLFALNQ
        mmeDR0119.1   NSPDAAHPVS ALATAHDKLD QAVATAYGWE WPLN...EDQ VLERLLALNL
          mmeDR2267   KNP....... .LRDAQERLD AAVSAAYGLP AGA......D MLDFLLALNA

                      1201                                                1250
      mmelfeLORF3P    KLTKKDW~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
        mmeLre121P    KLTKGER*~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
               mme    KMTSLLPPEK EIKKSKGKN* ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
          mmeNMA1791  KAAAIA~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
          mmeBSU0677  EMTER~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
             mmegcry  KLISHQP~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
            mmePflQ8  ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
             saro3834 KMTGGRSSEG GAA~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
          mmeMSI135   AREAFNEHAK SGLKTRKPRR RPTPEEVRRA PQMKLPIAGG RKSVVGPQQL
          mmeCC0826   ERAAAGR~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
        mmeDR0119.1   ERCPA~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
          mmeDR2267   XVAAAEARGA AVTGPGLPAG LNTADFVTAD AVRPLG*~~~ ~~~~~~~~~~

                      1251                1273
      mmelfeLORF3P    ~~~~~~~~~~ ~~~~~~~~~~ ~~~
        mmeLre121P    ~~~~~~~~~~ ~~~~~~~~~~ ~~~
               mme    ~~~~~~~~~~ ~~~~~~~~~~ ~~~
          mmeNMA1791  ~~~~~~~~~~ ~~~~~~~~~~ ~~~
          mmeBSU0677  ~~~~~~~~~~ ~~~~~~~~~~ ~~~
             mmegcry  ~~~~~~~~~~ ~~~~~~~~~~ ~~~
            mmePflQ8  ~~~~~~~~~~ ~~~~~~~~~~ ~~~
             saro3834 ~~~~~~~~~~ ~~~~~~~~~~ ~~~
          mmeMSI135   TTKDRENQPT SAERPRNTKR RTS
          mmeCC0826   ~~~~~~~~~~ ~~~~~~~~~~ ~~~
        mmeDR0119.1   ~~~~~~~~~~ ~~~~~~~~~~ ~~~
          mmeDR2267   ~~~~~~~~~~ ~~~~~~~~~~ ~~~
```